Some topics in the analysis of spherical data.

Thesis

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.21954/ou.ro.0000f900

SOME TOPICS IN THE ANALYSIS OF SPHERICAL DATA

by

Andrew T.A. Wood

B.A. (Cambridge), Dip. Math. Stat. (Cambridge)

A Thesis submitted for the Degree of

Doctor of Philosophy

in the Open University

May, 1985.

Author's Number: HDJ 63198

Date of Submission: May 1985

Date of Award: 8.8.85

Some Topics in the Analysis of Spherical Data

Ph.D. Thesis by Andrew T.A. Wood

ABSTRACT

This thesis is concerned with the statistical analysis of directions in 3 dimensions. An important reference is the book by Mardia (1972). At the time of publication of this book, the repertoire of spherical distributions used for modelling purposes was rather limited, and there was clearly a need to investigate other possibilities. In the last few years there has been some interest in the 8 parameter family of distributions mentioned by Mardia (1975), which is known as the Fisher-Bingham family.

In Chapter 1 an outline of the thesis is given. The Fisher-Bingham family is discussed in Chapter 2, and an effective method for calculating the normalising constant is presented. Attention is then focussed on an interesting 6 parameter subfamily, and a simple rule is given for classifying the distributions in this subfamily according to type (unimodal, bimodal, 'closed curve'). Estimation and inference are then discussed, and the Chapter is concluded with a numerical example.

In Chapter 3, the family of bimodal distributions presented in Wood (1982) is described. Other bimodal models are also mentioned briefly.

The problem of simulating Fisher-Bingham distributions

is considered in Chapter 4. Some inequalities are derived and then used to construct suitable envelopes so that an acceptance-rejection procedure can be used.

In Chapter 5, the robust estimation of concentration for a Fisher distribution is considered, and L-estimators of the type suggested by Fisher (1982) are investigated. It is shown that the best of these estimators have desirable all-round properties. Indications are also given as to how these ideas can be adapted to other contexts.

Possibilities for further research are mentioned in Chapter 6.

References

Fisher, N.I. (1982). Robust estimation of the concentration parameter of Fisher's distribution on the sphere. Appl. Statist., 31, 152-154.

Mardia, K.V. (1972). Statistics of Directional Data. London: Academic Press.

Mardia, K.V. (1975). Statistics of Directional Data (with discussion). J.R. Statist. Soc. B, 37, 349-393.

Wood, A. (1982). A bimodal distribution on the sphere. Appl. Statist., 31, 52-58.

Declaration

This thesis is the result of my own work and contains nothing which is the outcome of work done in collaboration.

Chapters 2, 3, 4 and 5 are each based on papers:

1. Some notes on the Fisher-Bingham distribution on the sphere. (Accepted by Communications in Statistics.)
2. A bimodal distribution on the sphere. (Applied Statistics, 1982.)
3. Simulation of spherical distributions in the Fisher-Bingham family. (Provisionally accepted by Communications in Statistics.)
4. Some robust estimators of the concentration parameter of a Fisher distribution. (Provisionally accepted by Biometrika.)

Acknowledgements

I am grateful to the referees for the papers mentioned above for a number of helpful criticisms. The general suggestions of the referee for 'Simulation of spherical distributions ...' were followed up carefully and led to substantial changes in the simulation method proposed in Chapter 4.

I have two very efficient typists to thank. Mrs. D. Hough typed earlier versions of most of the Chapters, and Mrs. M. Carter typed the final version of the thesis.

I am deeply grateful to my supervisor, Professor T. Lewis, for many illuminating discussions and for his support and encouragement throughout this project.

CONTENTS

CHAPTER 1

INTRODUCTION

In a number of scientific fields, including Geology, Meteorology, Astronomy and Biology, data consisting of directions or orientations often arise. There has been a corresponding need to develop statistical methods for analysing such data. In this thesis attention is limited to directions, as opposed to orientations; and the main concern is with directions in 3 dimensions (the spherical case). However, some of the results presented here also have implications for directional statistics in 2 dimensions (the circular case), and when this is so it is indicated in the text.

The book by Mardia (1972) provides an excellent account of the more theoretical aspects of directional statistics, and includes a number of applications to scientific problems. Some historical detail, and an extensive pre-1972 bibliography, are also given there. At the time of publication of Mardia's book, the repertoire of spherical distributions used for modelling purposes was rather limited, and there was clearly a need to investigate other possibilities. In the last few years there has been some interest in the 8 parameter family of distributions mentioned by Mardia (1975), which is known as the Fisher-Bingham family. This is a rich family of distributions, but there are a number of technical difficulties involved in its use. Interesting subfamilies have been discussed by Bingham and Mardia (1978) and Kent (1982).

In Chapter 2, the Fisher-Bingham family is discussed, and in particular an effective method for calculating the normalising constant in the general case is presented. Then attention is focussed on an interesting 6 parameter subfamily, and a relatively simple rule is given in terms of the parameters for classifying the distributions in the subfamily according to type (unimodal, bimodal or 'closed-curve'). Estimation and inference are discussed and, in the unimodal case, an approximate confidence cone for the modal direction is given. The Chapter is concluded with a numerical example.

In Chapter 3, a 5 parameter family of bimodal distributions is described. In this family, there is no constraint on the angle between the modal directions, but the modes are equal. Estimation and inference are discussed in some detail, and a practical example involving paleomagnetic data is presented.

We return to the Fisher-Bingham family in Chapter 4 and discuss how one might simulate these distributions. Some inequalities are derived, and are used to construct suitable envelopes so that an acceptance-rejection procedure can be used.

As in other areas of statistics, it is desirable to have robust estimation procedures available. In the case of the Fisher distribution, the maximum likelihood estimator of the mean direction is fairly robust against extreme observations, but it is generally acknowledged that the maximum likelihood estimator of concentration is not. In Chapter 5, we follow Fisher's (1982) suggestion and use certain L-statistics to estimate concentration. The asymptotic properties of these estimators are obtained, though there are a number of technical difficulties to be overcome in the derivation, and use needs to be made of Dudley's (1978) recent work on the convergence of empirical processes. It is shown, using Hampel's (1974) influence curve, that these estimators have desirable robustness properties; and the results of a simulation study indicate that the best of the L-statistic estimators have good small sample properties when the Fisher distribution is correct. Finally, suggestions are given as to how these ideas can be adapted to other contexts, for example to the robust estimation of the eigenvalues of the covariance matrix of a multivariate Normal distribution.

Possibilities for further research are indicated in Chapter 6.

CHAPTER 2

THE FISHER-BINGHAM FAMILY

2.1 Introduction

The Fisher-Bingham, or von Mises-Fisher-Bingham, family of distributions on the unit sphere, S_3, has densities which can be expressed in the form

$$\{c(\kappa, \mu, A)\}^{-1}\exp(\kappa x'\mu + x'Ax) \tag{2.1.1}$$

at a point $x \in S_3$. κ, a scalar, μ, a unit vector, and A, a 3x3 matrix, are the parameters. This family has been named so because each of the densities is proportional to the product of a Fisher density and a Bingham density.

Most distributions in current use in directional statistics possess a considerable amount of symmetry, for example rotational symmetry (Fisher distributions) or antipodean symmetry (Bingham distributions). While for many datasets such symmetry assumptions may be reasonable, it would seem desirable to have distributions available that are more suitable for data which do not exhibit any obvious symmetry. The Fisher-Bingham family is of interest from this point of view, because it contains distributions which possess no symmetry at all.

However, the difficulties involved in its use are correspondingly greater. One of the main practical problems is the calculation of the normalising constant, which is needed in the maximum likelihood estimation of the parameters. A practicable method of calculating this is outlined in section 2.2 and discussed in more detail in 2.5.

The Fisher-Bingham family was first mentioned in Mardia (1975), who observed that the maximum entropy distributions for given $E(x)$ and $E(xx')$ are the Fisher-Bingham distributions.

Beran (1979) proposed a hierarchy of families of directional distributions. In 3 dimensions, the p^{th} family in the hierarchy is that with distributions whose densities are proportional to $\exp\{\sum_{j=1}^{p} S_j(x)\}$, where each $S_j(x)$ is a linear combination of the spherical harmonics of order j . When $p = 1$, the Fisher family is obtained, and when $p = 2$, the Fisher-Bingham family. However, there appear to be certain theoretical and practical problems with the regression estimator that he proposes, which will be briefly mentioned in 2.2.

Kent (1982) has discussed the Fisher-Bingham family, with particular emphasis on a five-parameter subfamily which he suggests is a suitable analogue to the bivariate Normal on the plane. He shows that, for this subfamily, estimation and inference are tractable.

The Fisher-Bingham family on S_n , the unit hypersphere in n dimensions, is defined as in (2.1.1), but with μ and A replaced by n dimensional counterparts. In the main, only the three-dimensional case will be discussed here.

2.2 Calculation of the Normalising Constant

In this section, an alternative parametrisation for the Fisher-Bingham family is given. Using this alternative parametrisation it is readily seen that the normalising constant can be expressed as a relatively simple 1-dimensional integral. We need to prove the following straightforward result.

Lemma (2.2.1) Any Fisher-Bingham density f can be written in the form

$$f(x|\kappa,\rho,\mu,\mu_1,\mu_2) = c^{-1}\exp\{\kappa x'\mu+\rho(x'\mu_1)(x'\mu_2)\} \qquad (2.2.2)$$

where κ and ρ are scalars and μ , μ_1 and μ_2 are unit vectors, not in general orthogonal.

Note: the converse is obviously true.

Proof Since the elements of A in (2.1.1) only occur in the quadratic form $x'Ax$, we can without loss of generality take A to be symmetric. Also, since only unit vectors x are being considered, we can replace A with $A+\varepsilon I_3$, where ε is an arbitrary scalar and I_3 is the 3x3 identity matrix, without affecting the density in (2.1.1).

It follows that we can restrict attention to matrices

A of the form $Q'\lambda Q$, where Q is orthogonal and $\lambda = \mathrm{diag}\{0,\lambda_1,\lambda_2\}$ with $\lambda_1\lambda_2 \leqslant 0$. Now define

$$v_+ = (0,\sin\gamma,\cos\gamma) \quad \text{and} \quad v_- = (0,-\sin\gamma,\cos\gamma) .$$

Then if ρ and γ are chosen to be solutions of

$$-\rho\sin^2\gamma = \lambda_1 , \qquad \rho\cos^2\gamma = \lambda_2$$

and we put $\mu_1 = Q'v_+$ and $\mu_2 = Q'v_-$, it is easy to check that

$$\rho(x'\mu_1)(x'\mu_2) = x'Q'\lambda Qx = x'Ax \quad \text{for all} \quad x \in S_3 . \qquad \text{ooo}$$

We now make some comments:

i) The density of any Fisher-Bingham distribution on S_2, the unit circle, can be written as

$$c^{-1}\exp\{\kappa x'\mu_1+\rho(x'\mu_2)^2\}$$

where $x' = (\sin\theta,\cos\theta)$. In fact c, the normalising constant, has a convenient series expansion:

$$c = 2\pi\{\sum_{r=0}^{\infty} I_r(\kappa)I_{2r}(\rho)\cos^{-1}(2r\cos^{-1}(\mu_1'\mu_2))\}$$

I_r being the modified Bessel function of order r. Further details of the circular case are given in Yfantis and Borgman (1982).

ii) Not every Fisher-Bingham density on S_n, $n > 3$, can be parameterised as in (2.2.2): in general, additional terms in the exponent are needed. In fact, only densities for which the nxn matrix A in (2.1.1) has at most rank 3 have a representation of the form (2.2.2).

iii) Kent (1982) lists some of the important Fisher-Bingham subfamilies, defined by imposing simple constraints on the parameters. The corresponding constraints expressed in terms of the new parameters are as follows:

(a) Uniform: $\kappa = \rho = 0$

(b) Fisher: $\rho = 0$

(c) Dimroth-Watson: $\kappa = 0$, $\mu_1 = \pm \mu_2$

(d) Bingham: $\kappa = 0$

(e) FB_4: $\pm \mu = \pm \mu_1 = \pm \mu_2$

(f) FB_5: μ , μ_1, μ_2 mutually orthogonal.

iv) A 6-parameter subfamily of some interest, which we shall be discussing later, is obtained by imposing the constraint $\mu_1 = \pm \mu_2$, so that the exponent in (2.2.2) reduces to the sum of a Fisher and a Dimroth-Watson term. This subfamily, which we shall call FDW_6 , contains a fairly broad range of skew distributions. It has a 5-parameter subfamily, which we shall call FDW_5 , obtained by imposing the additional constraint $\mu'\mu_1 = \mu'\mu_2 = 0$. FDW_5 was mentioned by Kent in Barndorf-Neilson and Cox (1979, discussion).

We now use (2.2.2) to express the normalising constant, c , as a 1-dimensional integral. The axes can be chosen so that $\mu_1 = (0,0,1)$, $\mu_2 = (\sin 2\gamma, 0, \cos 2\gamma)$ and $\mu' = (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)$. Then, expressing x in polar coordinates, i.e. $x' = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, integrating with respect to ϕ , and using the substitution $u = \cos\theta$, we find that

$$c = 2\pi \int_{-1}^{1} \exp\{.\} I_0\{(1-u^2)^{\frac{1}{2}}[.]^{\frac{1}{2}}\}\,du \qquad (2.2.3)$$

where $\{ . \} = u\kappa\cos\alpha + u^2\rho\cos 2\gamma$

and $[.] = \kappa^2\sin^2\alpha + 2u\rho\kappa\sin\alpha\cos\beta\sin 2\gamma + u^2\rho^2\sin^2 2\gamma$,

I_0 being the modified Bessel function of degree 0 .

This integral can be speedily calculated with high relative accuracy using a suitable numerical integration procedure. Practical aspects are discussed in 2.5.

Two other methods for calculating the normalising constant will now briefly be mentioned.

(a) An obvious, but crude, way of obtaining c is to integrate the density (2.2.2) over the sphere numerically. However, when the log-likelihood is being maximised numerically, it is likely that the normalising constant will need to be evaluated at a large number of points in the parameter space. It would seem somewhat preferable to be required to evaluate a 1-dimensional, rather than a 2-dimensional, integral numerically a large number of times. On the whole, there should be a large reduction in the computer resources used.

(b) The normalising constant can be expressed as a series involving certain moments of a trivariate Normal distribution. This is a particular case of a result given in De Waal (1979) (that in which $p = 1$ and $m = 3$).

If, without loss of generality, we take A in (2.1.1) to be negative definite, and let $(-A)^{-\frac{1}{2}}$ be the (real) positive 'square root' of $(-A)^{-1}$, then, using Corollary 4.1 in De Waal (op.cit.)

$$c = 2 \sum_{r=0}^{\infty} Q_r / [(2r+1)!!r!] \tag{2.2.4}$$

where $Q_r = E\{(u+\frac{i\kappa}{\sqrt{2}}(-A)^{-\frac{1}{2}}\mu)'A(u+\frac{i\kappa}{\sqrt{2}}(-A)^{-\frac{1}{2}}\mu)\}^r$

and u is a zero-mean trivariate Normal with covariance the identity matrix, and $(2r+1)!! = 1.3....(2r+1)$.

De Waal notes that each Q_r can be expressed as a polynomial function of the cumulants $C_1,...,C_r$ (these polynomials are given in Kendall and Stuart (1963, p.69), though it would seem to be preferable to use the formulae given in Morris (1982, Theorem 2)) where each cumulant is given by the formula

$$C_j = 2^{j-1}(j-1)!tr(-A)^j - \kappa^2.2^{j-2}.j!\mu'(-A)^{j-1}\mu.$$

If either κ or the maximum of the differences between the eigenvalues of A , or both, are large (e.g. greater than 10), then a large number of terms may have to be included in (2.2.4) before convergence occurs. However, in view of the formulae in Morris (1982) this should not pose too much of a problem, and we would guess that, overall, there should not be much to choose between De Waal's method and the one suggested here.

We conclude this section by mentioning the method proposed in Beran (1979) for parameter estimation in the hierarchical families of directional distributions which he defines (recall that the Fisher-Bingham is one of these families). This estimator, which he calls a regression estimator, is based on a non-parametric ('window') density estimator of the true density. However, the validity of the asymptotic results given seem to be in doubt, because the crucial Lemma 1 in Beran (op.cit.) is wrong. In particular, statement 3.10 in the proof of Lemma 1 is invalid. (The incorrectness of Lemma 1 was pointed out by a referee for Wood (1982)).

A practical problem is that it is not clear how to go about choosing the sequence $\{c_n\}$ of positive numbers which are to converge to zero at a certain rate (see Beran (1979, section 3.2)). In our own experience, the regression estimate shows great sensitivity to the choice of $\{c_n\}$.

2.3 Fisher-Bingham densities : Stationary Points and Distribution Types in the General Case

Given a Fisher-Bingham density, for convenience parametrised as in (2.2.1), the number and type of its stationary points can be determined numerically : generally, it involves finding the real roots of a polynomial of degree 6.

Assume without loss of generality that the matrix A in (2.1.1) is of the form $A = \mathrm{diag}\{a_1, a_2, a_3\}$. Then, adding a Lagrangian term for the constraint $x'x = 1$, with Lagrangian multiplier λ , to the exponent in (2.1.1), differentiating with respect to x and equating to zero, we obtain

$$\kappa\mu + 2(A-\lambda I_3)x = 0 \qquad (2.3.1)$$

where I_3 is the identity matrix.

Now, a unit vector x will be a stationary point of the density if and only if there exists a λ such that λ and x satisfy (2.3.1). Suppose that λ and x are such a pair. Then two cases can arise:

(i) $\lambda = a_j$ $j=1,2$ or 3 and $A-\lambda I_3$ is singular, or

(ii) $\lambda \neq a_j$ $j=1,2$ or 3 and $A-\lambda I_3$ is non-singular.

In (ii), $x = -\kappa(A-\lambda I_3)^{-1}\mu/2$. Using the constraint $x'x = 1$, it is seen that

$$(\kappa^2/4)\sum_{j=1}^{3} \mu_j/(a_j-\lambda)^2 = 1$$

is satisfied or, equivalently

$$Q(\lambda)/\{(a_1-\lambda)(a_2-\lambda)(a_3-\lambda)\}^2 = 0$$

where $\mu' = (\mu_1, \mu_2, \mu_3)$ and

$$\begin{aligned} Q(\lambda) = {} & (a_1-\lambda)^2(a_2-\lambda)^2(a_3-\lambda)^2 \\ & -(\kappa^2/4)[\mu_1(a_2-\lambda)^2(a_3-\lambda)^2 + \mu_2(a_1-\lambda)^2(a_3-\lambda)^2 \\ & + \mu_3(a_1-\lambda)^2(a_2-\lambda)^2] \end{aligned} \qquad (2.3.2)$$

is a polynomial of degree 6.

To find the stationary points of the density, we solve the polynomial $Q(\lambda) = 0$. Then for any real solution $\lambda \neq a_j$, $x = -\kappa(A-\lambda I_3)^{-1}\mu/2$ will be a stationary point. In exceptional cases, there may be stationary points, possibly an infinite number (consider the subfamily FB_4 , which contains symmetric small circle distributions), corresponding to $\lambda = a_1, a_2$ or a_3 . These can easily be found by inspection of (2.3.1). The nature of each stationary point, x , is determined by the nature of $(A-\lambda I_3)$, restricted to the tangent plane at x .

There are three broad types of distribution in the Fisher-Bingham family: (i) unimodal, (ii) bimodal and (iii) 'closed curve' i.e. the probability is concentrated on a closed curve on S_3 .

2.4 Stationary Points and Distribution Types in the Fisher-Dimroth-Watson subfamily

The motivation for focussing attention on Fisher-Bingham subfamilies remains, because it appears that conditions, expressed in terms of the parameters, for a distribution in the full Fisher-Bingham family to be of a particular type would be very difficult to obtain. In this section we shall discuss what we have called FDW_6 , the Fisher-Dimroth-Watson subfamily, obtained by imposing the constraint $\mu_1 = \mu_2$ in (2.2.2). It contains a broad range of distributions: unimodal distributions, with and without rotational or elliptical symmetry; a wide range of bimodal distributions; and 'closed curve' (c.f. small circle) distributions.

For convenience, we shall use the following parametrisation:

$$f(x|\kappa,\rho,\alpha,\beta,\gamma,\delta) = c^{-1}\exp\{\kappa x'R\mu+\rho(x'\nu)^2\} \qquad (2.4.1)$$

with κ and ρ scalars, μ and ν unit vectors and R an orthogonal matrix, where

$$\mu' = (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha) ,$$

$$\nu' = (\sin\gamma\cos\delta, \sin\gamma\sin\delta, \cos\gamma)$$

and
$$R = \begin{bmatrix} \cos\gamma\cos\delta & -\sin\delta & \sin\gamma\cos\delta \\ \cos\gamma\sin\delta & \cos\delta & \sin\gamma\sin\delta \\ -\sin\gamma & 0 & \cos\gamma \end{bmatrix} .$$

The parameter space is defined by $\kappa \varepsilon [0,\infty)$, $\rho \varepsilon (-\infty,\infty)$, $\alpha \varepsilon [0,\pi/2]$, $\beta \varepsilon [0,2\pi)$, $\gamma \varepsilon [0,\pi]$ and $\delta \varepsilon [0,2\pi)$.

β , γ and δ are rotation parameters, (γ,δ) being the polar coordinates for the 'Dimroth-Watson' axis. α , which is the acute angle between the 'Fisher' direction and the 'Dimroth-Watson' axis, and ρ and κ , are shape parameters.

When $\alpha = 0$, FB_4 is obtained. This subfamily was discussed by Bingham and Mardia (1978), who were interested in it as a model for data concentrated about a small circle. When $\alpha = \pi/2$, we obtain the subfamily that we have named FDW_5 above. In fact, the part of this subfamily defined by $\alpha = \pi/2$ and $\rho \geq 0$ has the same qualitative features as FB_5; we would expect to be able to approximate closely any distribution in FB_5 with a distribution in FDW_5.

(Note: we can define the Fisher-Dimroth-Watson subfamily in n dimensions as the collection of distributions with densities of the form (2.4.1), but with μ, ν and R replaced by n dimensional counterparts. In 2 dimensions the Fisher-Bingham and Fisher-Dimroth-Watson subfamilies coincide. When $n > 3$, the normalising constant for the latter can be expressed as a 1-dimensional integral similar to that in (2.2.3), though with a different Bessel function.)

When $\beta = \gamma = 0$ in (2.4.1) and x is expressed in polar coordinates, the density takes a particularly simple form:

$$f(x|\kappa,\rho,\alpha) = c^{-1}\exp\{\kappa\sin\alpha\sin\theta\cos\phi+\kappa\cos\alpha\cos\theta+\rho\cos^2\theta\} \quad . \quad (2.4.2)$$

Our aim now is to determine which ranges of values of κ, ρ and α correspond to which qualitative types of distribution. We do this by determining how the number (and type) of stationary points of the density $f(x)$ depend on κ, ρ and α. Let

$$H(\theta,\phi|\kappa,\rho,\alpha) = \kappa\sin\alpha\sin\theta\cos\phi+\kappa\cos\alpha\cos\theta+\rho\cos^2\theta \quad . \quad (2.4.3)$$

Then

$$\partial H/\partial\theta = \kappa\sin\alpha\cos\theta\cos\phi-\kappa\cos\alpha\sin\theta-2\rho\sin\theta\cos\theta \qquad (2.4.4)$$

and $$\partial H/\partial\phi = -\kappa\sin\alpha\sin\theta\sin\phi \quad . \qquad (2.4.5)$$

Any point (θ,ϕ) is a stationary point if and only if $\partial H/\partial\theta$ and $\partial H/\partial\phi$ are zero there. It is clear from (2.4.5) that, unless $\kappa\sin\alpha = 0$, any stationary point can only occur on a great circle, which can be defined as $\{(\theta,\phi): \theta\in[0,2\pi), \phi = 0\}$.

When $\kappa\sin\alpha = 0$, (2.4.2) reduces to an FB_4 density. For FB_4 and FDW_5 (i.e. when $\alpha = 0$ or $\alpha = \pi/2$ respectively) our problem is easy to solve. We only mention here that in these two cases, the density will have precisely two stationary points (one a maximum, and one a minimum) if and only if $\kappa \geqslant 2|\rho|$.

Now suppose that $\kappa\sin\alpha > 0$. Then, putting $\phi = 0$, we substitute $t = \tan(\theta/2)$ and $\sigma = 2\rho/\kappa$ in (2.4.4), and after some manipulation it can be seen that $\partial H/\partial\theta = 0$ if and only if

$$\kappa(1+t^2)^{-2}[(\sin\alpha)t^4+2(\cos\alpha-\sigma)t^3+2(\cos\alpha+\sigma)t-\sin\alpha] = 0 \; . \qquad (2.4.6)$$

Since $\kappa\sin\alpha$ is assumed to be strictly positive, the term multiplying the brackets can be ignored, and the problem becomes that of determining how the number of real zeros of quartic equations of the form

$$P(t|\,\alpha,\sigma)=(\sin\alpha)t^4+2(\cos\alpha-\sigma)t^3+2(\cos\alpha+\sigma)t-\sin\alpha = 0 \qquad (2.4.7)$$

depends on α and σ . We give two results for the collection of quartics $\{P\} = \{P(t|\alpha,\sigma): \alpha \in (0,\pi/2), \sigma \in (-\infty,\infty)\}$.

<u>Proposition (2.4.8)</u> For any $P(t|\alpha,\sigma) \in \{P\}$, there is precisely one root in each of the intervals $(0,1)$ and $(-\infty,-1)$. Further, if $\sigma < 0$ these roots lie in $(\tan(\alpha/2),1)$ and $-\cot(\alpha/2),-1)$ respectively; if $\sigma = 0$ at $\tan(\alpha/2)$ and $-\cot(\alpha/2)$; and if $\sigma > 0$, in $(0,\tan(\alpha/2))$ and $(-\infty,-\cot(\alpha/2))$.

<u>Proposition (2.4.9)</u> There exist continuous curves $\Sigma_-(\alpha),\Sigma_+(\alpha)$ defined for $\alpha \in (0,\pi/2)$ such that if $\sigma \leqslant \Sigma_-(\alpha)$, then $P(t|\alpha,\sigma)$ has two roots in $(-1,0)$ and none in $(1,\infty)$; if $\sigma \geqslant \Sigma_+(\alpha)$, $P(t|\alpha,\sigma)$ has no roots in $(-1,0)$ and two roots in $(1,\infty)$; and if $\Sigma_-(\alpha) < \sigma < \Sigma_+(\alpha)$, then $P(t|\alpha,\sigma)$ has no roots in either $(-1,0)$ or $(1,\infty)$.

In the proof of these results, we shall make use of Budan's Theorem (see Dickson (1945)), which we state here.

<u>Budan's Theorem</u> Let $P(t)$ be a polynomial of degree n with real coefficients, and suppose c_1 and c_2 are real numbers. Then, if V_i is the number of variations in sign of the sequence $P(c_i)$, $P'(c_i)$, $P''(c_i) \ldots P^{(n)}(c_i)$ $(i=1,2)$, the number of roots between c_1 and c_2 is $|V_1-V_2|$, or less than this by a positive even integer. A root of multiplicity m is counted as m roots.

<u>Proof of (2.4.8)</u>

Let $G(\theta|\alpha,\sigma) = -\frac{1}{\kappa}\frac{\partial H}{\partial \theta}\Big|_{\phi=0} = \sin(\theta-\alpha)+(\sigma/2)\sin 2\theta$.

It is easy to show that

$$\text{(i)} \quad G(\pi/2-\theta|\alpha,\sigma) = -G(\theta|\pi/2-\alpha,-\sigma) \tag{2.4.10}$$

and
$$\text{(ii)} \quad G(\theta+\pi|\alpha,\sigma) = -G(\theta|\alpha,-\sigma) \quad . \tag{2.4.11}$$

It follows from (i) that $P(t|\alpha,\sigma)$ and $P(t|\pi/2-\alpha,-\sigma)$ have the same number of zeros in $(0,1)$. And from (ii), it is seen that $P(t|\alpha,\sigma)$ has the same number of zeros in $(0,1)$ as $P(t|\alpha,-\sigma)$ has in $(-\infty,-1)$. So to prove the first part of (2.4.8), we simply need to show that when $\sigma \leqslant 0$ and $\alpha \in (0,\pi/2)$, $P(t|\alpha,\sigma)$ has precisely one root in $(0,1)$. We do this using Budan's Theorem, taking $c_1 = 0$ and $c_2 = 1$. It is readily seen that when $\sigma \leqslant 0$ and $\alpha \in (0,\pi/2)$, $V_1 = 1$ and $V_2 = 0$. Hence there is precisely one root in each of the intervals $(0,1)$ and $(-\infty,-1)$.

To prove the second part of (2.4.8), we put

$$P(t|\alpha,\sigma) = Q(\alpha,t)+R(\sigma,t) \tag{2.4.12}$$

where $Q(\alpha,t) = (\sin\alpha)t^4+2(\cos\alpha)t^3+2(\cos\alpha)t-\sin\alpha$ and $R(\sigma,t) = 2\sigma t(1-t^2)$. It is not difficult to show that $Q(\alpha,t) < 0$ for $t \in (0,\tan(\alpha/2))$, $Q(\alpha,t) = 0$ at $t = \tan(\alpha/2)$, and $Q(\alpha,t) > 0$ for $t \in (\tan(\alpha/2),1)$. And clearly, for $t \in (0,1)$, $R(\sigma,t) > 0$ if $\sigma > 0$, $R(\sigma,t) = 0$ if $\sigma = 0$ and $R(\sigma,t) < 0$ if $\sigma < 0$.

So, when $\sigma = 0$ the root in $(0,1)$ is at $t = \tan(\alpha/2)$. And when $\sigma < 0$ and $t \in (0,\tan(\alpha/2)]$, $P(t|\alpha,\sigma) < 0$ since $Q(\alpha,t) \leqslant 0$ and $R(\sigma) < 0$. So in this case, the root must lie in $(\tan(\alpha/2),1)$. The corresponding assertions for $\sigma > 0$ and the interval $(-\infty,-1)$ are proved in a similar fashion. 000

Proof of (2.4.9) We can deduce the following from Budan's Theorem: when $\sigma \leq \cos\alpha + \sin\alpha$, there are no roots in $(1,\infty)$; and when $\sigma > \cos\alpha + \sin\alpha$, there are either no roots or 2 roots in $(1,\infty)$. (Put $c_1 = 1$, $c_2 = \infty$. Then $V_1 = 0$ if $\sigma \leq \sin\alpha + \cos\alpha$, and $V_1 = 2$ if $\sigma > \cos\alpha + \sin\alpha$; and $V_2 = 0$.)

We observe that for any $P(t|\alpha,\sigma) \varepsilon \{P\}$, $P(1|\alpha,\sigma) > 0$ and $P(\infty|\alpha,\sigma) > 0$. Using (2.4.12) it is clear that for any $t_0 \varepsilon (1,\infty)$ and $\alpha \varepsilon (0,\pi/2)$ we can find a σ_0 such that $P(t_0|\alpha,\sigma_0) < 0$ (take $\sigma_0 > Q(\alpha,t_0)/\{2t_0(t_0^2-1)\}$). Because of its continuity, $P(t|\alpha,\sigma_0)$ must then have one root in $(1,t_0)$ and one root in (t_0,∞). Also, for any $\varepsilon > 0$ $P(t_0|\alpha,\sigma_0+\varepsilon) < 0$ since $P(t_0|\alpha,\sigma_0+\varepsilon) = P(t_0|\alpha,\sigma_0)+R(\varepsilon,t_0)$ and both the latter terms are strictly negative. Therefore, if $P(t|\alpha,\sigma)$ has two roots in $(1,\infty)$, then, when $\varepsilon > 0$, so does $P(t|\alpha,\sigma+\varepsilon)$ and in the latter case they must be distinct.

For each $\alpha \varepsilon (0,\pi/2)$, define

$$\Sigma_+(\alpha) = \inf\{\sigma | P(t|\alpha,\sigma) \text{ has two roots in } (1,\infty)\}. \quad (2.4.13)$$

From above, $\Sigma_+(\alpha) \geq \sin\alpha + \cos\alpha > 0$. Clearly Σ_+ has the desired properties. To obtain $\Sigma_-(\alpha)$, we note that (2.4.11) implies that the number of roots $P(t|\alpha,\sigma)$ has in $(1,\infty)$ is the same as the number of roots $P(t|\alpha,-\sigma)$ has in $(-1,0)$. As a consequence $\Sigma_-(\alpha) = -\Sigma_+(\alpha)$. It also follows, from (2.4.10), that $\Sigma_-(\alpha) = -\Sigma_+(\pi/2-\alpha)$, so that $\Sigma_+(\alpha) = \Sigma_+(\pi/2-\alpha)$. 000

It is possible to determine $\Sigma(\alpha)$ $(=\Sigma_+(\alpha))$ analytically. We already know (from the proof of (2.4.9)) that if $|\sigma| > \Sigma(\alpha)$, then $P(t|\alpha,\sigma)$ can not have a double root. In fact, an argument by contradiction shows that when $\sigma = \pm\Sigma(\alpha)$, $P(t|\alpha,\sigma)$ does have a double root. So $P(t|\alpha,\sigma)\varepsilon\{P\}$ has a double root if and only if $\sigma = \pm\Sigma(\alpha)$.

It can be shown that if $P(t|\alpha,\sigma)$ has a double root at $t^* = \tan(\theta^*/2)$, then both $\partial H/\partial\theta$ and $\partial^2 H/\partial\theta^2$ are zero at $\theta = \theta^*$ and $\phi = 0$. So at any double root, α and σ must satisfy

$$-(\sigma/2)\sin(2\theta^*)-\sin(\theta^*-\alpha) = 0 \qquad (2.4.14)$$

and $$-\sigma\cos(2\theta^*)-\cos(\theta^*-\alpha) = 0\,. \qquad (2.4.15)$$

A little manipulation shows that (2.4.14) and (2.4.15) are simultaneously satisfied if and only if

$$[\tan^3\theta^*+\tan\alpha]/\{\tan\theta^*[\tan\theta^*-\tan\alpha]\} = 0\,. \qquad (2.4.16)$$

The only real solution of (2.4.16) is $\theta^* = \tan^{-1}(-(\tan\alpha)^{1/3})$. It follows that $\Sigma(\alpha) = 2\sin(\alpha-\theta^*)/\sin(2\theta^*)$. It is sensible to define the end points by $\Sigma(0) = \Sigma(\pi/2) = 1$.

All the discussion concerning stationary points of

Fisher-Dimroth-Watson distributions applies in any dimension. So, we have in fact extended the results given in Yfantis and Borgman (1982), whose interest is in the circular case.

It is a straightforward matter to determine the manner in which any $P(t|\alpha,\sigma)$ changes sign in the vicinity of its zeros (bearing in mind certain statements made in the proof of (2.4.9)). Using this information, we can determine the nature of the Hessian matrix for $H(\theta,\phi|\kappa,\rho,\alpha)$, at any stationary point; therefore, we can determine the nature of any stationary point.

We are now in a position to make the following statements for distributions of the form (2.4.2), when $\kappa > 0$ and $\alpha \in (0,\pi/2)$. We omit the cases $\kappa = 0$ (Dimroth-Watson), $\alpha = 0$ (FB_4) and $\alpha = \pi/2$ (FDW_5) for convenience, since the corresponding details are easily filled in. Firstly a word on nomenclature: if we say that a function $f(x,y)$ has an "A" - "B" stationary point at (x_o,y_o) we mean that $f(x,y_o)$ as a function of x has a type "A" stationary point at x_o, and $f(x_o,y)$ as a function of y has a type "B" stationary point at y_o.

(a) There is a mode at $(\theta,\phi) = (\theta_1,0)$, where if $\rho > 0$, then $\theta_1 \in (0,\alpha)$, if $\rho = 0$ then $\theta_1 = \alpha$, and $\rho < 0$ then $\theta_1 \in (\alpha,\pi/2)$.

(b) There is an antimode at $(\theta,\phi) = (\theta_2,\pi)$, where if $\rho > 0$, $\theta_2 \in (\pi-\alpha,\pi)$, if $\rho = 0$, $\theta_2 = \pi - \alpha$ and if $\rho < 0$, $\theta_2 \in (\pi/2,\pi-\alpha)$.

(c) If $2\rho > \kappa.\Sigma(\alpha)$, there is a min-max stationary point at $(\theta,\phi) = (\theta_3,0)$, and a mode at $(\theta_4,0)$, where $\pi/2 < \theta_3 < \theta_4 < \pi$. These distributions are of bimodal type, with modes at $(\theta_1,0)$ and $(\theta_4,0)$.

(d) If $-\kappa.\Sigma(\alpha) \leqslant 2\rho \leqslant \kappa.\Sigma(\alpha)$, there are no further stationary points unless $2\rho = \pm\kappa\Sigma(\alpha)$. If $2\rho = \kappa\Sigma(\alpha)$, there is an inflection-max stationary point at $(\theta,\phi) = (\theta_3,0)$, where $\theta_3 \in (\pi/2,\pi)$. And if $2\rho = -\kappa\Sigma(\alpha)$, there is an inflection-min stationary point at (θ_3,π) , where $\theta_3 \in (0,\pi/2)$. These distributions are of unimodal type, with mode at $(\theta_1,0)$.

(e) If $2\rho < -\kappa\Sigma(\alpha)$, there is an antimode at $(\theta,\phi) = (\theta_3,\pi)$ and a max-min stationary point at (θ_4,π) , where $0 < \theta_3 < \theta_4 < \pi/2$. These distributions are unevenly concentrated about the union of the two curves of steepest ascent, with respect to the density, from (θ_4,π) to $(\theta_1,0)$. In this sense, these are closed-curve distributions.

Note: When $\kappa\sin\alpha$ is greater than about 5, these distributions are practically indistinguishable from unimodal distributions, though for convenience we still refer to them as closed-curve distributions below.

To summarise: FDW_6 distributions fall into one of three categories: if $2\rho > \kappa.\Sigma(\alpha)$, they are of bimodal type; if $2|\rho| \leqslant \Sigma(\alpha)$, they are of unimodal type; and if $2\rho < -\kappa\Sigma(\alpha)$, they are of closed-curve type.

One of the interesting features of the FDW_6 subfamily is that it contains unimodal distributions with skew contours of constant density: that is, densities with no rotational or elliptical symmetry, though they do have 'half-elliptical' symmetry. The exact form of the dependence of the shape of the contours on κ , ρ and α seems to be complicated. However, we offer a rule-of-thumb, based on limited practical experience.

(i) When $0 < \alpha < \pi/2$ and $0 < 2\rho \leq \kappa . \Sigma(\alpha)$, the contours are like ellipses which have been distorted into a 2-dimensional egg-shape, but with symmetry about the major axis retained.

(ii) When $0 < \alpha < \pi/2$ and $-\kappa . \Sigma(\alpha) \leq 2\rho < 0$, the contours are again like distorted ellipses, but this time with symmetry about the minor axis retained.

2.5 Maximum Likelihood Estimation and some Hypothesis Tests

The full Fisher-Bingham family (FB_8) is a regular exponential family (in the sense of Barndorf-Nielson (1978)) and so, from the point of view of maximum likelihood estimation in particular, it will have some desirable properties. We state two of these:

(i) The log-likelihood is strictly concave in any natural parametrisation. Therefore, if the maximum likelihood estimates (MLE's) exist, they are unique.

(ii) If $x_1,\ldots,x_n$, where $n > 3$, is an independent sample from any spherical distribution which is absolutely continuous with respect to Lebesque measure, then the Fisher-Bingham MLE's exist with probability 1.

However, most of the subfamilies we have mentioned, in particular Dimroth-Watson, FB_4 , FB_5 , FDW_5 and FDW_6 , are curved exponential families. Geometrically, each of these subfamilies can be regarded as a curved hypersurface embedded in the natural parameter space of FB_8 (which is R^8) . In general, likelihood functions for curved exponential families may have more than one local maximum. So, since the aim of the maximum likelihood procedure is to find the global maximum of the likelihood, there is a problem: to establish whether or not MLE's for one of the subfamilies, obtained numerically, are the global MLE's.

A graphical procedure, which gives the necessary information about the number and approximate location of the local maxima of the likelihood for these curved exponential subfamilies, is suggested below, though it tends to require quite a large amount of computer resources. Firstly, we mention that the likelihood function for the Dimroth-Watson subfamily almost always has precisely 2 maxima: one in the region $\rho > 0$ and one in $\rho < 0$ (Mardia (1972, p.253)). For the other 4 subfamilies, although we have no theoretical results to offer, practical experience suggests that the likelihood functions for FB_4 and FB_5 generally only have one maximum; but for FDW_5 and FDW_6 two maxima:

one in the region $\rho > 0$ and one in $\rho < 0$.

The graphical procedure is based upon the following simple observation: for each fixed γ and δ, (2.4.1) is a 4-parameter exponential family, which satisfies (i) and (ii) above (though obviously these families are not of statistical interest because they are not closed under rotations).

For a sample $x_1, \ldots, x_n$ of unit vectors, the log-likelihood based on (2.4.1) is:

$$L_n(\kappa, \rho, \alpha, \beta, \gamma, \delta) = \sum_{j=1}^{n} \log\{f(x_j | \kappa, \rho, \alpha, \beta, \gamma, \delta)\} . \quad (2.5.1)$$

Then for each fixed γ and δ, L_n maximised over κ, ρ, α and β has a unique maximum, $L^*_n(\gamma, \delta)$ say, at

$$\kappa^* = \kappa^*(\gamma, \delta), \rho^* = \rho^*(\gamma, \delta), \alpha^* = \alpha^*(\gamma, \delta), \beta^* = \beta^*(\gamma, \delta) . \quad (2.5.2)$$

Clearly, L^*_n has a local maximum at a point (γ, δ) if L_n has one there: so all the required information concerning the local maxima of L_n is contained in L^*_n. The function L^*_n can be computed over a grid of points numerically, and its contours, which yield the required information concerning the local maxima, can be plotted using a graphics routine.

We now mention some practical aspects of the procedure we suggested in 2.2 for obtaining the MLE's. To maximise the log-likelihood function, we have used routine E04JAF from the NAG library (see NAG), a routine which does not require specification of the functional form of any of the partial derivatives. This routine allows one to impose simple constraints on the variables of the function being maximised. So, by imposing appropriate constraints, we can fit any of the subfamilies we have mentioned, using essentially the same procedure. However, when fitting FB_8, it appears to be better to use a natural parametrisation, though for the subfamilies, the parametrisation (2.2.2) is adequate.

At each point in the parameter space at which the log-likelihood is evaluated, the normalising constant must be calculated. In practice, it is crucial to bear in mind that we only need be concerned with relative accuracy, and not absolute accuracy, in the following sense: if c is the true value of the integral, and c^* is an estimate of c such that $c = c^*(1+\varepsilon)$, then we only require that ε, the relative error, be small, and we do not need to worry about the magnitude of $|c-c^*|$, the absolute error, because we only need to calculate $\log c$.

The integration routine we have used, D01AJF from NAG, allows one to choose which aspect of the error is to be controlled, and it can also cope with integrands with sharp peaks, a possibility that may arise when the parameters are large. Since each time the integration routine is called, the integrand is calculated at a number of points, it is important that the modified Bessel function, I_0, be calculated efficiently. This can be done using routine S18AEF (NAG), in which I_0 is expressed as a sum of Chebyshev functions, and different expansions are used for different ranges of values of the argument.

We make two further practical points. Firstly, it is better to maximise the log-likelihood without imposing restrictions on the permitted ranges of the parameters, and then to choose the maximum likelihood estimates such that they lie in suitable ranges afterwards, in order to avoid 'solutions' which lie on an artificial boundary.

Secondly, the question of initial estimates. If a natural parameterisation is used for FB_8, the choice of initial estimates does not seem to significantly affect the performance of the procedure. For FDW_5 and FDW_6, initial estimates can be obtained from the graphical procedure mentioned above. If the graphical procedure is not used, then a number of different initial estimates, perhaps based on the Fisher or Dimroth-Watson MLE's, should be tried.

The procedure has worked well, even when the maximum values of the arguments of I_0 and the exponential in (2.2.2) approach 150, approximately their maximum permitted values on the computer we have used.

Kent (1982) mentions some hypothesis tests for the Fisher-Bingham family. Also, an omnibus goodness-of-fit test for the Fisher distribution, against an unspecified

Fisher-Bingham alternative, is given in Mardia, Holmes and Kent (1984). These tests are based on 'score' statistics. Computationally, such tests tend to be simpler than the corresponding likelihood ratio tests, though in general they have less power asymptotically.

We mention a sequence of likelihood ratio tests suitable for data which we have reason to believe are generated from a unimodal distribution. The reason(s) may be due to prior knowledge, or based on the outcome of a test of unimodality (see 3.5), or the result of a graphical study of the data. See Lewis and Fisher (1982) for a discussion of some graphical methods for spherical data. Tests appropriate for bimodal models and closed-curve models are mentioned in the next Chapter.

Below, $L(H_j)$ will denote the globally maximised log-likelihood under hypothesis H_j.

(a) $H_o : FDW_6$ versus $H_1 : FB_8$.

Under H_o, $-2[L(H_o)-L(H_1)]$ is asymptotically χ^2_2. Reject H_o if this statistic is too large. In the unimodal case, this can be interpreted as a test of whether the contours have a 'line' of symmetry, as opposed to no symmetry.

(b) $H_o : FDW_5$ versus $H_1 : FDW_6$.

Under H_o, $-2[L(H_o)-L(H_1)]$ is asymptotically χ^2_1; reject H_o if this statistic is too large. Interpretation: a test of elliptical symmetry versus symmetry about only one line.

(c) H_o:Fisher versus H_1:FDW_5 .

Under H_o , $-2[L(H_o)-L(H_1)]$ is asymptotically χ^2_2 ; reject H_o if it is too large. Interpretation: a test of rotational symmetry versus elliptical symmetry.

In 2.7 the usefulness of these tests is discussed. It appears that very large samples are required if test (a) is to be effective.

It may sometimes be desirable to include FB_4 in this sequence of tests, as it is a more general rotationally symmetric model than the Fisher distribution. The extent to which FB_4 is a useful unimodal, rotationally symmetric generalisation of the Fisher distribution is not yet clear (see section 6.1).

2.6 A confidence Region for the Modal Direction

Suppose we are given an FDW_6 distribution $f(x|\kappa,\rho,\alpha,\beta,\gamma,\delta)$ parametrised as in (2.4.1) with $0<\alpha<\pi/2$. If ψ is the acute angle between the modal direction and the 'Dimroth-Watson' axis, then the modal direction, λ say, is given by $\lambda = R\zeta$, where $\zeta = (\sin\psi\cos\beta, \sin\psi\sin\beta, \cos\psi)$ and R is as in (2.4.1).

Straightforward computation yields

$$\lambda = \begin{bmatrix} \sin\psi\cos\beta\cos\gamma\cos\delta - \sin\psi\sin\beta\sin\delta + \cos\psi\sin\gamma\cos\delta \\ \sin\psi\cos\gamma\sin\delta + \sin\psi\sin\beta\cos\delta + \cos\psi\sin\gamma\sin\delta \\ -\sin\psi\cos\beta\sin\gamma + \cos\psi\cos\gamma \end{bmatrix}. \qquad (2.6.1)$$

Now suppose we are given a random sample of n unit vectors from the above distribution, and that we have obtained maximum likelihood estimates $\hat{\kappa}, \hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}$ and $\hat{\delta}$ under the FDW_6 hypothesis. Then an approximate $(100-\varepsilon)\%$ confidence region for λ, the modal direction is given by

$$\{x \mid x \varepsilon S_3, (x'\hat{\omega}_1)^2/\hat{V}_1 + (x'\hat{\omega}_2)^2/\hat{V}_2 \leq t_\varepsilon/n\} \qquad (2.6.2)$$

where: $\hat{\omega}_1$ is the unit vector in the direction $\hat{\lambda} - \hat{\nu}/\cos\hat{\psi}$, $\hat{\omega}_2$ is the unit vector in the direction of the vector product of $\hat{\lambda}$ and $\hat{\nu}$, $\hat{\lambda} x \hat{\nu}$; ν is as in (2.4.1), and the MLE's of λ, ν and ψ are defined implicitly; and t_ε is given by $Pr(\chi^2_2 > t_\varepsilon) = \varepsilon$. $\hat{V}_1$ and $\hat{V}_2$ are estimates of positive numbers V_1 and V_2 determined below.

The derivation of (2.6.2) is essentially very straightforward, if a little involved, so we shall only give an outline.

(i) From the standard asymptotic theory for MLE's,

(a) the MLE's are consistent,

i.e. $(\hat{\kappa}, \hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}) \to (\kappa, \rho, \alpha, \beta, \gamma, \delta)$ in probability.

(b) $\sqrt{n}[(\hat{\kappa}, \hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}) - (\kappa, \rho, \alpha, \beta, \gamma, \delta)] \to N(0, I^{-1})$ in distribution, where $I = I(\kappa, \rho, \alpha, \beta, \gamma, \delta)$ is the information matrix, for a sample size of 1, evaluated at the true parameter values.

(ii) For computational convenience, we choose a coordinate system such that the true values of the rotation parameters, in the new coordinate system, are $\beta = \delta = 0$ and $\gamma = \pi/2$.

(iii) In this new coordinate system we expand the MLE of λ, $\hat{\lambda}$, in a Taylor series about the true modal direction and obtain

$$\hat{\lambda} = \lambda + A(\hat{\kappa}-\kappa, \hat{\rho}-\rho, \hat{\alpha}-\alpha, \hat{\beta}-\beta, \hat{\gamma}-\gamma, \hat{\delta}-\delta)' + o_p(1) \qquad (2.6.3)$$

where the 3x6 matrix A is given by

$$\begin{bmatrix} -\sin\psi \dfrac{\partial\psi}{\partial\kappa} & -\sin\psi \dfrac{\partial\psi}{\partial\rho} & -\sin\psi \dfrac{\partial\psi}{\partial\alpha} & -\sin\psi & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\psi & \sin\psi \\ -\cos\psi \dfrac{\partial\psi}{\partial\kappa} & -\cos\psi \dfrac{\partial\psi}{\partial\rho} & -\cos\psi \dfrac{\partial\psi}{\partial\alpha} & -\cos(\psi) & 0 & 0 \end{bmatrix}$$

and the derivatives in A are evaluated at the true parameter values. In fact (2.6.2) can be written

$$\sqrt{n}(\hat{\lambda}-\lambda)' = -(\sin\psi, 0, \cos\psi).Z_1 + (0,1,0).Z_2 \qquad (2.6.4)$$

where $Z_1 = \sqrt{n}(\hat{\kappa}-\kappa)\dfrac{\partial\psi}{\partial\kappa} + \sqrt{n}(\hat{\rho}-\rho)\dfrac{\partial\psi}{\partial\rho} + \sqrt{n}(\hat{\alpha}-\alpha)\dfrac{\partial\psi}{\partial\alpha} + \sqrt{n}(\hat{\gamma}-\gamma)$

and $Z_2 = \sqrt{n}(\hat{\delta}-\delta)\cos\psi + \sqrt{n}(\hat{\beta}-\beta)\sin\psi$.

$\hat{\lambda}$, and the two vectors on the RHS of (2.6.4), are three mutually orthogonal unit vectors.

(iv) We can obtain expression for $\partial\psi/\partial\kappa$, $\partial\psi/\partial\rho$

and $\partial\psi/\partial\alpha$ by differentiating (2.4.7) implicitly, where $t = \tan(\psi/2)$.

(v) It follows from (i) that Z_1 and Z_2 are asymptotically Normal. When $\gamma = \pi/2$ and $\beta = \delta = 0$, it is easy to see that the information matrix takes the form

$$I = \begin{bmatrix} J & 0 \\ 0 & K \end{bmatrix} \qquad (2.6.5)$$

where $J = J(\kappa,\rho,\alpha,\gamma)$ is 4x4 and $K = K(\delta,\beta)$ is 2x2 . As a consequence, Z_1 and Z_2 are asymptotically independent.

(vi) $Z_1 \to N(0,V_1)$ and $Z_2 \to N(0,V_2)$, where $V_1 = g'J^{-1}g$ $g = (\partial\psi/\partial\kappa, \partial\psi/\partial\rho, \partial\psi/\partial\alpha, 1)$, and J as in (2.6.5), evaluated at the true parameter values; and $V_2 = (\sin\psi,\cos\psi)K^{-1}(\cos\psi,\sin\psi)'$, with K as in (2.6.5). Clearly V_1 and V_2 do not depend on the coordinate system chosen.

(vii) Asymptotically, Z_1 can be interpreted as the angular error of the estimate along the great circle in the plane defined by the Fisher and Dimroth-Watson axes (i.e. along ω_1) , and Z_2 as the angular error in the perpendicular direction (i.e. along ω_2) . Since $(Z_1{}^2/V_1 + Z_2{}^2/V_2) \to \chi^2_2$, (2.6.2) is appropriate as a confidence region for the modal direction.

(viii) In practice we shall need to estimate V_1 ,

V_2 and ω_1 , ω_2 . We can do this by replacing the true parameter values with their MLE's in the appropriate formulae.

2.7 Some Numerical Results

Two Fisher-Bingham distributions, which we shall call D_1 and D_2 were selected such that, in parametrisation (2.1.1),

(i) D_1 has parameters $\kappa = 4$, $\mu' = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ and $A = \mathrm{diag}\{0,3,6\}$ and

(ii) D_2 has parameters $\kappa = 8$, $\mu' = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ and $A = \mathrm{diag}\{0,6,12\}$.

Using the method outlined in 2.3, it was shown that, in both cases $Q(\lambda)$ has only two real roots. So D_1 and D_2 have only two stationary points, a maximum and a minimum, and are therefore unimodal distributions. These particular distributions were chosen as they might be expected, in each case, to be among the more asymmetric unimodal FB_8 distributions of similar degrees of concentration.

The first two moments $t_j = E(x|D_j)$ and $S_j = E(xx'|D_j)$ of D_j $(j=1,2)$ were calculated by suitably adjusting the integrand in (2.2.3). Then treating (t_1, S_1) and (t_2, S_2) as though they were the first two sample moments, obtained from independent, identically distributed samples, FB_8 , FDW_6 , FDW_5 and Fisher distributions were fitted to each, by maximum likelihood. Of course, this can be done because the first two sample moments are sufficient statistics for the FB_8 family. A selection of the results are presented in Tables 2.1 and 2.2.

Put $\Omega_j(\kappa,\mu,A) = -\log(c/2\pi)+\kappa t_j{}'\mu+\mathrm{tr}(AS_j)$. $(j=1,2)$

Then, if (t_j,S_j) had been obtained from a sample size of n , $n\Omega_j$ would be the log-likelihood function. Now, in Table 2.1 Λ_1 is Ω_1 evaluated at the true parameter values, and $\Lambda(FB_8)$, $\Lambda(FDW_6)$, $\Lambda(FDW_5)$ and $\Lambda(\text{Fisher})$ are the global maxima, obtained using the numerical procedure described earlier, of Ω_1 over the FB_8 , FDW_6 , FDW_5 and Fisher hypotheses respectively. The integer-valued function m is defined as follows: $m(FB_8,FDW_6,0.05)$, for example, is the smallest sample size such that, given sample moments (t_1,S_1) , the hypothesis FDW_6 would be rejected in favour of FB_8 , at the 95% level, using the standard large sample likelihood ratio test. (In other words, m is the smallest integer that that $2m\,[\Lambda(FB_8)-\Lambda(FDW_6)] > \alpha$ where $\Pr(\chi^2_2 > \alpha) = 0.05$. And so on.

SK_1 and SK_2 are non-parametric measures of skewness about the mean direction which are described later.

We note the closeness of the true values of the normalised likelihood Λ_1 and Λ_2 and the estimated values $\Lambda(FB_8)$ in Tables 2.1 and 2.2. In both cases, the initial

TABLE 2.1

$D_1 : \kappa = 4, \mu' = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ $A = \text{diag}\{0,3,6\}$

modal direction: $(0.16, 0.274, 0.948)'$

$t_1' = (0.207, 0.32, 0.7)$ $S_1 = \begin{bmatrix} 0.119 & 0.063 & 0.123 \\ 0.063 & 0.230 & 0.169 \\ 0.123 & 0.169 & 0.651 \end{bmatrix}$

$(t_1't_1)^{1/2} = 0.796$ $SK_1 = 0.07$ $SK_2 = 0.199$

$\Lambda_1 = 0.8570$ $\Lambda(FB_8) = 0.8570$

$\Lambda(FDW_6) = 0.8223$ $\Lambda(FDW_5) = 0.7335$

$m(FB_8, FDW_6, 0.05) = 86$ $m(FDW_6, FDW_5, 0.05) = 22$

Note: Definitions are given on pages 33 and 34.

TABLE 2.2

D_2 : $\kappa = 8, \mu' = (1/\sqrt{3},\ 1/\sqrt{3},\ 1/\sqrt{3})$ $A = \text{diag}\{0,6,12\}$

modal direction: $(0.16,\ 0.274,\ 0.948)'$

$t_2' = (0.197,\ 0.303,\ 0.863)$ $S_2 = \begin{bmatrix} 0.075 & 0.059 & 0.161 \\ 0.059 & 0.156 & 0.237 \\ 0.161 & 0.237 & 0.769 \end{bmatrix}$

$(t_2't_2)^{1/2} = 0.935$ $SK_1 = 0.034$ $SK_2 = 0.075$

$\Lambda_2 = 1.8548$ $\Lambda(FB_8) = 1.8548$

$\Lambda(FDW_6) = 1.8341$ $\Lambda(FDW_5) = 1.8148$

$m(FB_8, FDW_6, 0.05) = 144$ $m(FDW_6, FDW_5, 0.05) = 100$

Note: Definitions are given on pages 33 and 34.

estimates were some distance from their final values, so that the maximisation procedure has clearly worked well. The final parameter estimates were also accurate to at least 4 significant figures. In both 2.1 and 2.2 the likelihood functions for FDW_5 and FDW_6 had two maxima. In all cases the global maximum was chosen.

In so far as D_1 and D_2 are good representatives of the more asymmetric unimodal FB_8 distributions with similar degrees of concentration, these results seem to suggest the following: that for unimodal data, if the concentration about the mode is moderate, departures from elliptical symmetry can be detected, when present, using the likelihood ratio test, even in reasonably small samples (e.g. in Table 2.1, $m(FDW_6, FDW_5, 0.05) = 22$) . But for data sufficiently highly concentrated about a point, very large samples will be required to detect such departures (e.g. in Table 2.2, $m(FDW_6, FDW_5, 0.05) = 100$) .

We shall now describe the measures of skewness SK_1 and SK_2 . Suppose that F is a distribution on S_3 . If we express the points $x \in S_3$ in polar coordinates, (θ, ϕ) , relative to a frame of reference defined by μ_1 , μ_2 and μ_3 , three mutually orthogonal unit vectors, so that

$$x = \sin\theta\cos\phi\mu_1 + \sin\theta\sin\phi\mu_2 + \cos\theta\mu_3$$

then SK_1 and SK_2 are defined by

$$SK_1 = SK_1(\mu_3,F) = \{[E(\cos\phi)]^2+[E(\sin\phi)]^2\}^{\frac{1}{2}}$$

and $$SK_2 = SK_2(\mu_3,F) = E\{[E(\cos\phi|\theta)]^2+[E(\sin\phi|\theta)]^2\}^{\frac{1}{2}} .$$

SK_1 and SK_2 do not depend on the choice of μ_1 and μ_2. It is easily shown that $0 \leqslant SK_1(\mu_3,F) \leqslant SK_2(\mu_3,F) \leqslant 1$.

A meaningful choice for μ_3 is the median direction of F. In this case, SK_1 and SK_2 are zero for any Fisher, Bingham, FB_5 or FDW_5 distribution. SK_1 is of more practical interest than SK_2 because it can be defined for sample $x_1,\ldots,x_n$ of unit vectors. Since $SK_1 = 0$ when μ_3 is the median direction, the point for which the sum of the arc lengths to $x_1,\ldots,x_n$ is a minimum, when μ_3 is the mean direction SK_1 is a measure of skewness based on the discrepancy between the mean and median directions.

The fact that SK_1 and SK_2 are smaller in Table 2.2 indicates that the contours of the densities (which are the same for D_1 and D_2) sufficiently close to the mode are only mildly skew. Of course, this is to be expected on theoretical grounds, because of the smoothness of FB_8 densities.

We conclude this section with a vague conjecture: that asymmetries will generally be easier to detect in data from bimodal or closed-curve distributions, than in the unimodal case. This is because for unimodal data, especially highly concentrated unimodal data, the mean direction will be 'close' to one of the eigenvectors of the matrix of second moments, but for bimodal and closed-curve data, this need not be so to the same degree.

CHAPTER 3

BIMODAL MODELS

3.1 Introduction

In the analysis of directional data, there are occasions when a bimodal model is required. Schmidt (1976) gives an example of spherical data which seem to be of a bimodal nature.

We shall be focussing attention on two bimodal models: one is the family of distributions described in Wood (1982) and the other is the Fisher-Dimroth-Watson family, described in the previous chapter.

We begin by briefly mentioning some bimodal models on the circle. Mardia and Spurr (1973) discuss m-modal distributions (where m is an integer) with densities of the form

$$\{2\pi I_0(\kappa)\}^{-1}\exp\{\kappa\cos(m\theta-\mu_0))\} \qquad (3.1.1)$$

obtained by rescaling von Mises distributions. The modes occur at $\mu_0 + 2\pi r/m$ $(r=0,\ldots,m=1)$ and are of equal magnitude. When $m = 2$ (the bimodal case) the modes are 180^o apart.

Another model, the 'discrete mixture' model, is given by mixtures of two von Mises distributions. Conditions under which such mixtures are bimodal, as opposed to unimodal, are given in Mardia and Sutton (1975). However, in the general (5-parameter) case, inference proves to be awkward. Marida (1972, p.71) discusses the special case in which the modes are 180° apart; in this case, these difficulties do not arise.

A third possibility is to use the Fisher-Bingham family on the circle, which contains both symmetric and asymmetric bimodal distributions; though there do not appear, as yet, to be examples of the use of this bimodal model in the literature.

The spherical models we shall be considering here can be viewed as analogues of the first and third circular models mentioned above. To obtain the first, we "double the longitude" of a Fisher distribution (analogous to "doubling the angle" of a von Mises distribution). The spherical construction gives bimodal distributions with equal mode strengths (as does the circular construction), and distributions with modal directions having any angular separation can be obtained (unlike the circular case, for which they will always be 180° apart).

The other bimodal model that we shall be discussing, the Fisher-Dimroth-Watson family, can be regarded as an analogue of the Fisher-Bingham distribution on the circle (for the reasons given in the previous chapter). It contains bimodal distributions with modal directions of any angular separation, and also with differing mode magnitudes.

We omit discussion of the 'discrete mixture' model on the sphere. In the general (7-parameter) case there are difficulties in inference corresponding to those on the circle, but more severe. And rather than consider simplifications of the mixture model, we have decided to concentrate attention on the two models mentioned above.

3.2 A 5-parameter family of Bimodal Distributions

Consider the distribution on the unit sphere, S_3, which is given by

$$f'(\theta,\phi|\alpha,\beta,\kappa)dS=c(\kappa)\exp\{\kappa\cos\alpha\cos\theta+\kappa\sin\alpha\sin\theta\cos(2\phi-\beta)\}\,dS.. \tag{3.2.1}$$

where (θ,ϕ) are the usual polar coordinates for a point on S_3, $dS = \sin\theta d\theta d\phi$, and $c(\kappa) = \kappa/\{2\pi(e^{\kappa}-e^{-\kappa})\}$. Denote the family of distributions of this form by D'. The density f' resembles the density of a Fisher distribution whose modal direction is (α,β), the difference being that $\cos(2\phi-\beta)$ replaces $\cos(\phi-\beta)$. In fact, the marginal density of θ for the distribution in (3.2.1) is the same as that of θ for a Fisher distribution whose modal direction is (α,β).

We note the following properties of the distributions in D'.

(i) When $\kappa = 0$, we obtain the uniform distribution.

In (ii) - (ix) it is supposed that $\kappa > 0$.

(ii) When $\alpha > 0$, the modes occur at $(\theta,\phi) = (\alpha, \beta/2)$ and $(\theta,\phi) = (\alpha, (\beta+2\pi/2))$. They are of equal magnitude. The anti-modes occur at $(\theta,\phi) = (\pi-\alpha, (\beta+\pi)/2)$ and $(\theta,\phi) = (\pi-\alpha, (\beta+3\pi)/2)$. These are also of equal magnitude.

(iii) The angle between the modal directions (and anti-modal directions) is 2α.

(iv) When $\alpha > 0$ it follows from (ii) that the projections of the two modal directions onto the X-Y plane make angles $\beta/2$ and $(\beta+2\pi)/2$ with the X axis. Hence β is a rotation parameter.

(v) κ is a concentration parameter - it specifies the concentrations about the modal directions (the larger κ, the larger the concentration).

(vi) When $\alpha = 0$, we obtain the Fisher distribution $F((0,0,1)',\kappa)$.

(vii) The effective parameter space can be taken to be: $0 \leq \alpha \leq \pi/2$, $0 \leq \beta < 2\pi$ and $\kappa \geq 0$.

(viii) The mean direction is $(0,0,1)'$, i.e. $Ex = \lambda(0,0,1)'$ where $x = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)'$ and $\lambda = \lambda(\alpha,\kappa) \geq 0$. In fact, it is straightforward to show that $\lambda = B(\kappa)\cos\alpha$ where $B(\kappa) = \coth(\kappa)-\kappa^{-1}$. So $\lambda = 0$ if and only if either $\kappa = 0$ or $\alpha = \pi/2$.

(ix) For the Fisher distribution, the contours of constant density are small circles, i.e. the sets $\{(\theta,\phi): \cos\alpha\cos\theta+\sin\alpha\sin\theta\cos(\phi-\beta)=\text{const.}\}$ are small circles on S_3. For f' in (3.2.1), the corresponding sets are given by $C_\varepsilon = \{(\theta,\phi): \cos\alpha\cos\theta+\sin\alpha\sin\theta\cos(2\phi-\beta)=\cos\varepsilon\}$. When $\varepsilon<\alpha$, C_ε will be the union of two non-intersecting closed curves on S_3, each one surrounding a mode.

The unit vectors $e_1 = (1,0,0)'$, $e_2 = (0,1,0)'$ and $e_3 = (0,0,1)'$ play a special role in distributions of the form (3.2.1). If we put $x = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)'$, then $x'e_1 = \sin\theta\cos\phi$, $x'e_2 = \sin\theta\sin\phi$ and $x'e_3 = \cos\theta$. To obtain the form of f' after an arbitrary rotation of axes, we generalise from e_1, e_2 and e_3 to three mutually orthogonal unit vectors
$\mu_1 = (\cos\gamma\cos\delta, \cos\gamma\sin\delta, -\sin\gamma)'$, $\mu_2 = (-\sin\delta, \cos\delta, 0)'$ and $\mu_3 = (\sin\gamma\cos\delta, \sin\gamma\sin\delta, \cos\gamma)'$, say, where $0\leqslant\gamma\leqslant\pi$ and $0\leqslant\delta<2\pi$ (only two parameters, γ and δ, are required here as β plays the role of the third rotation parameter). Using standard trigonometric formulae, we obtain

$$
\begin{aligned}
f(x\,|\,\gamma,\delta,\alpha,\beta,\kappa) = c(\kappa)\exp\Bigg\{ & \kappa\cos\alpha\,(x'\mu_3) + \\
& + \kappa\sin\alpha\cos\beta\left[\frac{(x'\mu_1)^2-(x'\mu_2)^2}{(1-(x'\mu_3)^2)^{\frac{1}{2}}}\right] + \\
& + \kappa\sin\alpha\sin\beta\left[\frac{2(x'\mu_1)(x'\mu_2)}{(1-(x'\mu_3)^2)^{\frac{1}{2}}}\right]\Bigg\}
\end{aligned}
$$

where f is the general form of f'. Also, we replace D' with D. Note that (γ,δ) are the polar coordinates for the mean direction, μ_3.

3.3 Estimation and the Information Matrix. Hypotheses Testing

Consider a sample of n unit vectors $x_1,\ldots,x_n$. Then the log-likelihood function with respect to D, L_n say, is given by

$$L_n(\gamma,\delta,\alpha,\beta,\kappa) = -n\log 2\pi - nA(\kappa) + \kappa(u\cos\alpha + (v\cos\beta + w\sin\beta)\sin\alpha) \tag{3.2.3}$$

where $u(\gamma,\delta) = \sum_{i=1}^{n} (x_i'\mu_3)$,

$$v(\gamma,\delta) = \sum_{i=1}^{n} \{[(x_i'\mu_1)^2-(x_i'\mu_2)^2]/[1-(x_i'\mu_3)^2]^{\frac{1}{2}}\},$$

$$w(\gamma,\delta) = \sum_{i=1}^{n} \{2(x_i'\mu_1)(x_i'\mu_2)/[1-(x_i'\mu_3)^2]^{\frac{1}{2}}\} \text{ and}$$

$$A(\kappa) = \log[(e^{\kappa}-e^{-\kappa})/\kappa].$$

A word on the notation in this section: for each (γ,δ), $\alpha^*(\gamma,\delta)$ will denote the MLE of α calculated as though (γ,δ) were the known true mean direction. Similarly for $\beta^*(\gamma,\delta)$ and $\kappa^*(\gamma,\delta)$. '^' will be used for the full MLE's. It is clear that $\hat{\alpha} = \alpha^*(\hat{\gamma},\hat{\delta})$,

$\hat{\beta} = \beta^*(\hat{\gamma}, \hat{\delta})$ and $\hat{\kappa} = \kappa^*(\hat{\gamma}, \hat{\delta})$.

Equating $\partial L_n/\partial\alpha$, $\partial L_n/\partial\beta$ and $\partial L_n/\partial\kappa$ to 0 , we obtain

$$\tan\alpha^* = (v^2+w^2)^{\frac{1}{2}}/u \ , \quad \alpha^* = \alpha^*(\gamma,\delta) \ ;$$

$$\tan\beta^* = w/v \ , \quad \beta^* = \beta^*(\gamma,\delta) \tag{3.2.4}$$

and $B(\kappa^*) = (u^2+v^2+w^2)^{\frac{1}{2}}/n$, $\kappa^* = \kappa^*(\gamma,\delta)$, where the function $B(.)$ is defined in 3.2,(viii).

We now consider the function

$$L_n^*(\gamma,\delta) = L_n(\gamma,\delta,\alpha^*,\beta^*,\kappa^*) \ . \tag{3.2.5}$$

Clearly $L_n^* \geq L_n$ and $\sup L_n^* = \sup L_n$. Using (3.2.3) and (3.2.4) L_n^* can be obtained explicitly. In fact, it takes the same form as for the Fisher distribution:

$$L_n^*(\gamma,\delta) = -n\log 2\pi + n(\kappa^* B(\kappa^*) - A(\kappa^*)) \ . \tag{3.2.6}$$

The RHS of (3.2.6) is an increasing function of κ^* ; and $B(\kappa^*)$ is an increasing function of κ^* . So finding (γ,δ) to maximise L_n^* is equivalent to finding (γ,δ) to maximise $B(\kappa^*) = (u^2+v^2+w^2)^{\frac{1}{2}}/n$, or, equivalently, $(u^2+v^2+w^2)$. That is, to find (γ,δ) we only need to maximise $(u^2+v^2+w^2)$, which can be done using a numerical maximisation procedure. We have found it a little easier to use a procedure which does not require specification of the functional form of the partial derivatives (e.g. subroutine E04JAF of the NAG library).

The question of an initial estimate for (γ, δ) remains. The sample mean direction, (γ_R, δ_R), a consistent estimator of (γ, δ), would appear to be the most sensible initial estimate, at least when the length of the sample resultant is not close to zero. However, when the "true" modal directions are widely separated, that is, α is close to $\pi/2$, it may be that the length of the sample resultant is close to zero. When this is the case, it may be better to obtain the eigenvector, t_2, corresponding to the middle eigenvalue of the matrix of sums of products of the sample (see Mardia (1972, equation (8.4.16)) and then to use the procedure twice: in one case, taking the direction of t_2 as the initial estimate, and the other, the direction of $-t_2$. Then the final estimate of (γ, δ) which gives the larger value of $(u^2+v^2+w^2)$ should be taken as $(\hat{\gamma}, \hat{\delta})$.

Let $I = \{E(-\partial^2 L/\partial\psi_i \partial\psi_j)\}$ be the information matrix, where $(\psi_1, \ldots, \psi_5) = (\gamma, \delta, \alpha, \beta, \alpha)$ and L is the log-likelihood for a sample size of 1. Then straightforward calculation gives

$$I = \begin{bmatrix} & & & 0 & 0 \\ & A & & 0 & 0 \\ & & & 0 & 0 \\ 0 & 0 & 0 & \kappa B(\kappa) & 0 \\ 0 & 0 & 0 & 0 & B'(\kappa) \end{bmatrix} \qquad (3.2.7)$$

where A is 3x3 matrix, with element $A(3,3) = E(-\partial^2 L/\partial\beta^2) = \kappa B(\kappa)\sin^2\alpha$. As functions of x, $\partial L/\partial\gamma$ and $\partial L/\partial\delta$ are bounded, and continuous except at

$x = \pm \mu_3$; and $\partial L/\partial\alpha$, $\partial L/\partial\beta$ and $\partial L/\partial\kappa$ are bounded and continuous. However, three of the second partial derivatives, namely $\partial^2 L/\partial\gamma^2$, $\partial^2 L/\partial\delta^2$ and $\partial^2 L/\partial\gamma\partial\delta$, are infinite at $x = \pm \mu_3$, when $\alpha > 0$. Nevertheless, $E(\partial^2 L/\partial\gamma^2)$, $E(\partial^2 L/\partial\delta^2)$ and $E(\partial^2 L/\partial\gamma\partial\delta)$ all exist. This can be shown by straightforward, but laborious, calculation.

The important point is that there is sufficient regularity for the asymptotic normality of the MLE's to hold when I is non-singular, and in this case they have covariance matrix I^{-1} . In fact, I is non-singular as long as $\kappa > 0$, $0 < \alpha \lesssim \pi/2$ and $0 < \gamma < \pi$. However, when $\alpha = 0$, α and β are not asymptotically normal. When I is non-singular, it follows from the form of I^{-1} that $\hat{\kappa}$ and $\hat{\alpha}$ will be aymptotically independent of each other, and the other three estimates $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\delta}$. (Of course, in this paragraph we have implicitly assumed that the MLE's and I are based on observations generated from a distribution in D).

There are several hypotheses concerning this bimodal family which could be of interest. Four are mentioned below, and large sample tests of these hypotheses against the more general alternative, D , are suggested.

For any hypothesis K , L_K will denote the maximum of the likelihood function when K is assumed. In cases (i)-(iv) below, the alternative hypothesis, G , is that the observations are from a distribution in D with no restrictions on the parameters (other than that they lie in the effective parameter space).

(i) Null hypothesis H1: the modal directions are v_1 and v_2, where v_1 and v_2 are prescribed unit vectors.

Specifying v_1 and v_2 is equivalent to specifying α, β, γ and δ; so L_{H1}, and hence the test statistics $\lambda_1 = -2\log[L_{H1}/L_G]$ should be easy to obtain. Under H1, λ_1 is asymptotically χ^2_4 (chi-squared, with four degrees of freedom); and H1 should be rejected if λ_1 is "large" compared with a χ^2_4 random variable.

(ii) Null hypothesis H2: $\alpha = \alpha_o$, where $\alpha_o (>0)$ is specified.

Asymptotically, α is independent of the other MLE's when G is true. When n is large, $n^{\frac{1}{2}}(\hat{\alpha}-\alpha_o)$ is approximately $N(0,(\kappa B(\kappa))^{-1})$ if H2 is true. H2 should be rejected if $|\hat{\alpha}-\alpha_o| > M/(n\kappa B(\kappa))^{\frac{1}{2}}$, where M satisfies $P(|N(0,1)| > M> = \varepsilon$, and ε is the chosen size of the test. This test is asymptotically equivalent to the corresponding likelihood ratio test.

(iii) Null hypothesis H3: $\alpha = 0$ (i.e. the Fisher hypothesis).

L_{H3} is the maximised likelihood under the Fisher hypothesis. To check the degrees of freedom, it is helpful to reparameterise by putting $L_1 = \sin\alpha\cos\beta$ and $L_2 = \sin\alpha\sin\beta$. Then the Fisher hypothesis is equivalent to H3: $L_1 = L_2 = 0$, and points in the parameter space corresponding to Fisher distributions are in the interior. H3 should be rejected if $\lambda_3 = -2\log[L_{H3}/L_G]$ is "large" compared with a χ^2_2 random variable.

In fact, a simpler test statistic is available, whose calculation does not involve numerical maximization. In the notation of (3.2.5), we define

$$L_1^*(\gamma,\delta) = \sin\alpha^*(\gamma,\delta)\cos\beta^*(\gamma,\delta) \quad \text{and}$$

$$L_2^*(\gamma,\delta) = \sin\alpha^*(\gamma,\delta)\sin\beta^*(\gamma,\delta) \; .$$

Then, when the Fisher hypothesis is true (but not otherwise):

$$\lambda'_3 = n\kappa^*(\gamma_R,\delta_R)B(\kappa^*(\gamma_R,\delta_R))\left[(L_1^*(\gamma_R,\delta_R))^2+(L_2^*(\gamma_R,\delta_R))^2\right]\to\lambda_3 \,,$$

in probability, where $B(.)$ is defined in 3.2(viii), and (γ_R,δ_R) is the direction of the sample resultant. Results of this nature can be found in Cox and Hinkley (1974, pp. 323-324).

The hypothesis H3 is likely to be of interest mainly when the true value of α is small; and when α is small the power of the test based on λ'_3 should not be substantially less than the power of the full likelihood ratio test.

(iv) Null hypothesis H4: $\alpha_o \leq \alpha \leq \pi/2$ (or $0 < \alpha \leq \alpha_o$) where $\alpha_o > 0$ is specified.

The recommended test for H4 is similar to that for H2 , except that a one-sided, rather than two-sided, test should be used.

3.4 Further Points

i) It was noted earlier that the distribution in the D family have equal mode strengths. In fact, each of these distributions has reflective symmetry about two perpendicular planes. Frequently, however, there will be no a priori reason for supposing that the underlying distribution (assumed bimodal) from which a given set of data is generated possesses such symmetry. So, given a set of data, assumed to be of a bimodal nature, we may ask: to what extent are these symmetry assumptions justified?

We shall not attempt a full discussion of this problem, though we do propose two simple non-parametric tests which should give useful indications.

Suppose $x_1,\ldots,x_n$ are independent identically distributed unit vectors from some bimodal distribution G, with density g, the theoretical modal directions of G being v_1 and v_2, say. Define

$S^1 = \{x : x \varepsilon S_3$ and $x'v_1 \geq x'v_2\}$ and

$S^2 = \{x : x \varepsilon S_3$ and $x'v_1 \leq x'v_2\}$. The S^1 and S^2 are complementary closed unit hemispheres. We define marginal densities on S^j $(j=1,2)$ by $h_j(u) = \int_{A(j,u)} g(x)dx$ $(j=1,2)$, where $A(j,u) = \{x : x \varepsilon S^j, x'v_j = u\}$, with corresponding distribution H_1 and H_2.

We now consider the null hypothesis:

$$K: P(x \varepsilon S^1) = P(x \varepsilon S^2) \text{ and } H_1 \equiv H_2 . \qquad (3.4.1)$$

(The dependence of (3.4.1) on v_1 and v_2 has been suppressed.) Any bimodal distribution with modal directions v_1 and v_2, and the double reflective symmetry mentioned above, will satisfy this null hypothesis. In particular, it will be satisfied by any distribution from D with modal directions v_1 and v_2. However, it seems that it is possible to construct bimodal distributions with modal directions v_1 and v_2 which satisfy the null hypothesis but do not have this double reflective symmetry; but, since only quick diagnostic tests are being proposed, we shall not pursue this point further.

Put $r = \sum_{j=1}^{n} u(x_j)$, where $u(x) = \begin{matrix} 0 & \text{if} & x \varepsilon S^1 \\ 1 & \text{if} & x \varepsilon S^2 \end{matrix}$.

Then, under the null hypotheses, $r \sim Bi(\frac{1}{2}, n)$ i.e. r is binomially distributed with parameters $\frac{1}{2}$ and n. Under any alternative, $r \sim Bi(p,n)$ for some $p \varepsilon [0,1]$. Using suitable estimates of v_1 and v_2, such as their maximum likelihood estimates under D, we can test the hypothesis that $p = \frac{1}{2}$ (i.e. that $P(x \varepsilon S^1) = P(x \varepsilon S^2)$) using r, in standard fashion. If this hypothesis is firmly rejected, then we should infer that the double symmetry assumption is inappropriate.

The second test is a Wilcoxon two sample test, conditional on the value of r, of whether $H_1 \equiv H_2$. Suppose, without loss of generality, that $x_1,\ldots,x_{n-r} \in S^1$ and $x_{n-r+1},\ldots,x_n \in S^2$. Define $s_j = x_j'v_1$ $(j=1,\ldots n-r)$ and $t_j = x_j'v_2$ $(j=n-r+1,\ldots n)$. Then the s's are i.i.d. from H_1 and the t's are i.i.d. from H_2. A version of the Wilcoxon statistic for two samples (conditional on the value of r) is given by

$$w = \sum_{i=1}^{r} \sum_{j=1}^{n-r} \operatorname{sgn}(s_i - t_j) \text{, where}$$

$\operatorname{sgn}(a) = 1$ (-1) if $a > 0$ $(a < 0)$. Under the null hypothesis, w is asymptotically Normal, with mean 0 and variance $r(n-r)(n+1)/3$. So, again using suitable estimates for v_1 and v_2, the null hypothesis, and therefore the double symmetry assumption, should be rejected if $|w|.\{3/(r(n-r)(n+1))\}^{\frac{1}{2}} > \Phi(1-\alpha/2)$ where Φ is the standard normal distribution function and α is the chosen size of the test.

ii) Consider the density on S_3 given by

$$f(\theta,\phi,m) = c(\kappa)\exp\{\kappa\cos\alpha\cos\theta + \kappa\sin\alpha\sin\theta\cos(m\phi-\beta)\} .$$

When $m = 2$, we obtain a member of D. For general m, where m is a positive integer, this will be the density of the m-modal distribution in which the modes are equally spaced on the small circle $\theta = \alpha$. It is straightforward to obtain the general form of these distributions, with the functions v and w of 3.3 suitably re-defined. Of course, as m increases, v and w become correspondingly more complicated, and the model more specialised. Nevertheless, the results given in 3.3 for the case $m = 2$ hold for any integer greater than 2.

iii) In principle, the Dimroth-Watson distributions can be extended in the same way that the Fisher distributions have been here. However, neither the maximum likelihood estimation of the parameters, nor the information matrix, simplify as in the Fisher case.

iv) Bimodal distributions on S_n, $n > 3$, can be constructed by 'doubling the longitude' of von Mises-Fisher distributions in analogous fashion.

Any point $x \varepsilon S_n$ can be expressed in polar coordinates as follows:

$$x' = (\sin\theta_1 \ldots \sin\theta_{n-2}\cos\phi, \sin\theta_1 \ldots \sin\theta_{n-2}\sin\phi, \\ \sin\theta_1 \ldots \sin\theta_{n-3}\cos\theta_{n-2}, \ldots, \sin\theta_1\cos\theta_2, \cos\theta_1)$$

where $\theta_1, \ldots, \theta_{n-2} \varepsilon [0,\pi]$ and $\phi \varepsilon [0,2\pi)$. von Mises-Fisher densities, expressed in polar coordinates are of the form

$$c_n(\kappa)\exp\{\kappa x'\mu\}\sin^{n-2}\theta_1, \ldots, \sin\theta_{n-2} \qquad (3.4.2)$$

where x is as above and μ is an n-dimensional unit vector.

If we 'double the longitude', i.e. replace ϕ by 2ϕ in (3.4.2), then the normalising constant remains unchanged and, analogously to the 3-dimensional case, $\theta_1,\ldots,\theta_{n-2}$ have the same marginal distribution as $\theta_1,\ldots,\theta_{n-2}$ from the corresponding von Mises-Fisher distribution.

After an arbitrary rotation of axes, and expressed in Cartesian coordinates, these distributions are of the form

$$f(x) = c_n(\kappa)\exp\{\kappa\cos\alpha(x'v_3)+\ldots$$

$$\ldots+\kappa\sin\alpha\left[(x'v_1)^2-(x'v_2)^2\right]/\left[(x'v_1)^2+(x'v_2)^2\right]^{\frac{1}{2}}$$

where v_1 , v_2 and v_3 are any 3 orthogonal unit vectors. There are $(3n-4)$ parameters. The modal directions are $\cos\alpha\ v_3 \pm \sin\alpha\ v_1$.

3.5 <u>The Fisher-Dimroth-Watson family as a Bimodal Model</u>

In 2.4, it was shown that a Fisher-Dimroth-Watson distribution is bimodal if and only if $2\rho > \kappa.\Sigma(\alpha)$, where κ , ρ and α are defined in (2.3.1) and the curve $\Sigma(.)$ is obtained from (2.4.16). In the special cases $\alpha = 0$, $\alpha = \pi/4$ and $\alpha = \pi/2$, this condition reduces to $2\rho > \kappa$, $\rho > \kappa$ and $2\rho > \kappa$ respectively.

With choice of axes as in (2.3.2) we have shown that the modes, expressed in polar coordinates, occur at $(\theta_1,0)$, where $0 \leqslant \theta_1 \leqslant \alpha$ and $(\theta_4,0)$ where $\pi/2 < \theta_4 \leqslant \pi$

(we include here the possibilities that $\kappa = 0$ and $\alpha = 0$).

Intuition should suggest that the value of the density (2.3.2) at the mode $(\theta_1, 0)$ is greater than (or equal to) that at the mode $(\theta_4, 0)$, and in fact it is easy to show that this is the case.

Proposition (3.5.1) If $2\rho > \kappa . \Sigma(\alpha)$, then
(i) $H[\theta_1, 0 | \kappa, \rho\alpha] \geq H[\theta_4, 0 | \kappa, \rho, \alpha]$ where H is defined in (2.3.6) and (ii) $\theta_4 \leq \pi - \theta_1$.

Proof

(i) $H[\theta, 0] - H[\pi - \theta, 0] = \kappa(\cos(\theta - \alpha) - \cos(\theta + \alpha)) = 2\kappa \cos\theta \cos\alpha$
which is non-negative for $\theta \in [0, \pi/2]$.
So, since $(\pi - \theta_4) \in [0, \pi/2]$ we have:

$$H[\theta_1, 0] = \sup_{\theta \in [0, \pi/2]} H[\theta, 0] \geq H[\pi - \theta_4, 0] \geq H[\theta_4, 0] .$$

(ii) It follows from results in 2.3 that:

$$\frac{\partial H}{\partial \theta}[\theta, 0] \geq 0 \quad \text{when} \quad 0 \leq \theta \leq \theta_1$$

$$\frac{\partial H}{\partial \theta}[\theta, 0] \leq 0 \quad \text{when} \quad \theta_1 \leq \theta \leq \pi/2$$

$$\frac{\partial H}{\partial \theta}[\pi - \theta, 0] \geq 0 \quad \text{when} \quad 0 \leq \theta \leq (\pi - \theta_4)$$

and $$\frac{\partial H}{\partial \theta}[\pi - \theta, 0] \leq 0 \quad \text{when} \quad (\pi - \theta_4) \leq \theta \leq \pi/2 .$$

But from (i), $\frac{\partial H}{\partial \theta}[\theta, 0] - \frac{\partial H}{\partial \theta}(\pi - \theta, 0] = -2\kappa \sin\theta \cos\alpha \leq 0$ for $\theta \in [0, \pi/2]$. As a consequence $\theta_4 \leq \pi - \theta_1$. Equality occurs only when either $\kappa = 0$, $\alpha = 0$ or $\alpha = \pi/2$.

We now mention some statistics and hypothesis tests of interest. Suppose we are in the situation of 2.6: that the MLE's under FDW_6, $\hat{\kappa}, \hat{\rho}, \hat{\alpha}, \hat{\gamma}, \hat{\delta}, \hat{\beta}$, have been obtained, and that the true parameter values are $\kappa, \rho, \alpha, \beta, \gamma$ and δ. It would clearly be desirable to have a statistic which distinguishes between unimodality and bimodality. In view of the results given in 2.4, a suitable statistic, based on the MLE's of the shape parameters would be

$$\hat{\tau} = \tau(\hat{\kappa}, \hat{\rho}, \hat{\alpha}) = (\hat{\kappa} . \Sigma(\hat{\alpha})/2\hat{\rho})^{-1} . \tag{3.5.2}$$

For bimodal distributions, $\tau(\kappa, \rho, \alpha) > 1$, for unimodal distributions $-1 \leqslant \tau(\kappa, \rho, \alpha) \leqslant 1$, and for closed curve distributions $\tau(\kappa, \rho, \alpha) < -1$. Then, the hypotheses of interest would be:

$$H_1 : \tau > 1 \qquad H_2 : -1 \leqslant \tau \leqslant 1 \qquad H_3 : \tau < -1 .$$

If $\hat{\tau} > 1$, we may wish to assess the extent to which this value is compatible with the underlying distribution being unimodal. To do this, we could test H_2 (null hypothesis) against H_1; similarly, when $-1 \leqslant \tau \leqslant 1$ we could test the null hypothesis H_1 (or H_3) against H_2, using the asymptotic normality of $\hat{\tau}$. Asymptotically, $\sqrt{n}(\hat{\tau} - \tau(\kappa, \rho, \alpha))$ is $N(0, V)$, where

$$V = L'I^{-1}L \quad \text{and} \quad L' = \left(\frac{\partial\tau}{\partial\kappa}, \frac{\partial\tau}{\partial\rho}, \frac{\partial\tau}{\partial\alpha}\right)$$

and $I = I(\kappa, \rho . \alpha)$ the information matrix for a sample of size 1, are both evaluated at the true parameter values.

In practice it will be necessary to estimate V, for example with its MLE, $\hat{V}$.

Formally, the tests are as follows, though probably the real feature of interest will be the significance levels. For H_2 versus H_1 : reject H_2 if $\sqrt{n}(\hat{\tau}-1) > C\hat{V}^{\frac{1}{2}}$, where C satisfies $P(N(0,\hat{V}) < C) = \alpha_C$, α_C being the chosen size of the test. And for H_1 versus H_2 : reject H_1 if $\sqrt{n}(\hat{\tau}-1) < C\hat{V}^{\frac{1}{2}}$, where here $P(N(0,\hat{V}) < C) = \alpha_C$. The corresponding details for the closed curve hypothesis are similar.

A test of the double symmetry assumptiom, discussed in 3.4, is readily available in the FDW_6 bimodal model. In view of proposition (3.5.1), the double symmetry assumption is satisfied if and only if $\kappa\cos\alpha = 0$. For this, we can use a standard likelihood ratio test; the resulting likelihood ratio statistic is χ^2_1 under the hypothesis that $\kappa\cos\alpha = 0$. So we should reject the double symmetry assumption if this statistic is too large.

An estimate of the angular distance between the modal directions, based on the MLE's is given by

$$2\left[\tan^{-1}(t_4(\hat{\kappa},\hat{\rho},\hat{\alpha}))-\tan^{-1}(t_1(\hat{\kappa},\hat{\rho},\hat{\alpha}))\right] \qquad (3.5.3)$$

where t_1 and t_4 are, respectively, the smallest and largest positive roots of the quartic equation $P(t|\alpha,\sigma)$ (see 2.3.10). To test hypotheses concerning the angle between the modal directions, such as the second and fourth hypotheses mentioned in 3.3 for the D family, or to obtain an approximate confidence interval for this angle, one can use the asymptotic normality of the statistic (3.5.3). It has asymptotic mean

$$2\left[\tan^{-1}(t_4(\kappa,\rho,\alpha))-\tan^{-1}(t_1(\kappa,\rho,\alpha))\right]$$

and variance given by

$$4L'I^{-1}L/n$$

where

$$L' = (1+t_4^2)^{-1}\left(\frac{\partial t_4}{\partial\kappa},\frac{\partial t_4}{\partial\rho},\frac{\partial t_4}{\partial\alpha}\right)-(1+t_1^2)^{-1}\left(\frac{\partial t_1}{\partial\kappa},\frac{\partial t_1}{\partial\rho},\frac{\partial t_1}{\partial\alpha}\right)$$

As in similar cases above, $I = I(\kappa,\rho,\alpha)$ and L should in theory be evaluated at the true parameter values; but in practice, they will need to be estimated. The terms $\frac{\partial t_4}{\partial\kappa}$ (etc) can be estimated by implicitly differentiating $P(t\,|\alpha,\sigma)$, rearranging, and then substituting $t = t_4(\hat\kappa,\hat\rho,\hat\alpha)$, $\alpha = \hat\alpha$ and $\sigma = 2\hat\rho/\hat\kappa$ (etc).

We conclude this section by noting that to test whether two prescribed unit vectors are the modal directions (the first hypothesis given in 3.3) is a rather awkward problem in the (full) FDW_6 model. This is because the modal directions depend on all 6 parameters, in a fairly complicated way. In the context of the parameter space, any such hypothesis corresponds to a (complicated) 4-dimensional hypersurface, embedded in a 6-dimensional rectangle. However, the details are somewhat simpler when double symmetry is assumed (i.e. under the constraint $\kappa\cos\alpha = 0$) .

3.6 A Palaeomagnetic Example

Table 3.1 gives a set of estimates of a previous magnetic pole position (Table 2 in Schmidt, 1976) obtained using palaeomagnetic techniques. Each estimate is associated with a different site, the 33 sites being spread over a large area of Tasmania. A plot of the data is given in Figure 3.1.

Schmidt shows that a Fisher model is not appropriate (using the test described by Watson and Irving, 1957) and says that "the data [i.e. pole estimates] appear to fall into two main groups which are derived from two distinct geographical regions". But since it is not known a priori how these two regions should be defined, it is not clear how the data should be separated into two groups. This indicates the need for a bimodal model.

An analysis using D was performed. The length and direction of the sample resultant were found to be 30.76 and $(37.7^\circ, 155.7^\circ)$ respectively. The MLE's were obtained numerically, using NAG routine E04JAF, to a satisfactory approximation:

$(\gamma,\delta) = (37.7^\circ, 152.5^\circ)$, $\alpha = 10.7^\circ$, $\beta = 178^\circ$ and $\kappa = 19.3$.

TABLE 3.1

Estimates of a previous magnetic pole position of the Earth

	θ^o	ϕ^o		θ^o	ϕ^o
1.	34.3	148.3	18.	42.2	159.6
2.	42.8	176.5	19.	37.4	164.8
3.	26.2	162.2	20.	27.0	168.6
4.	40.9	196.5	21.	48.5	193.6
5.	38.0	147.3	22.	38.1	164.4
6.	33.3	178.5	23.	50.7	138.5
7.	59.8	183.9	24.	41.1	107.8
8.	56.5	184.1	25.	40.7	173.3
9.	67.7	109.4	26.	45.9	170.8
10.	47.1	133.4	27.	36.7	174.4
11.	42.7	125.9	28.	38.9	174.6
12.	34.4	124.4	29.	44.6	144.7
13.	37.5	148.5	30.	49.5	125.6
14.	44.3	75.2	31.	49.2	173.5
15.	40.5	142.0	32.	36.7	165.7
16.	30.5	94.9	33.	48.1	187.8
17.	21.3	209.2			

$(90-\theta)$ is the latitude (S^o) and ϕ is the longitude (E^o)

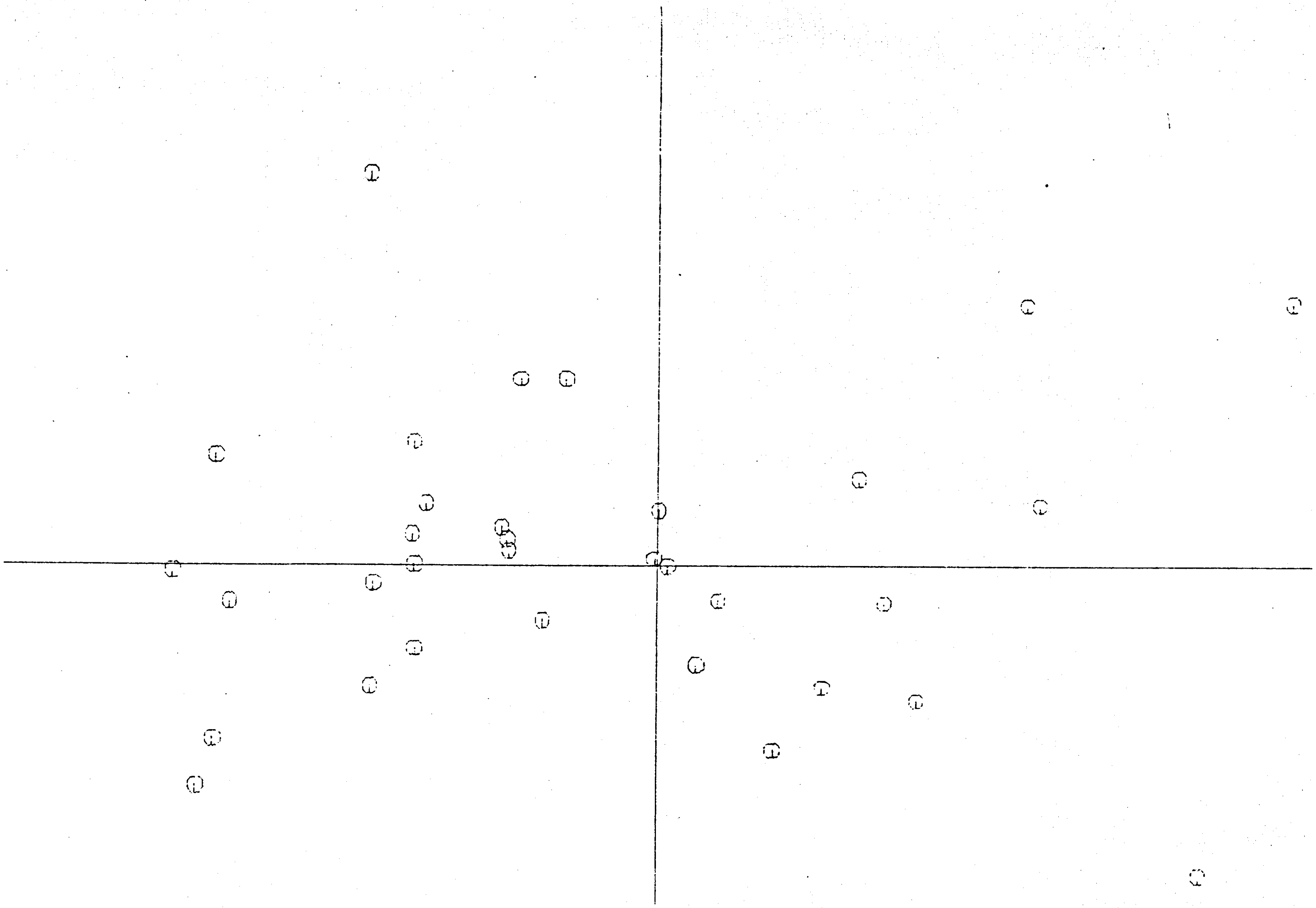

FIGURE 3.1. A projection of the data onto the plane using Lambert's equal-area projection, as described in Selby (1964).

λ_3 , the statistic for testing H_3 : Fisher versus G : the bimodal alternative D , was calculated as 17.8. Under H_3 , λ_3 has χ^2_2 distribution for large n . Hence the conclusion is that the Fisher hypothesis should be rejected strongly, since $P(\chi^2_2 > 10.597) = 0.005$. λ'_3 , the alternative test statistic given, was found to be 19.37. So the alternative test rejects the Fisher hypothesis even more emphatically.

Confidence intervals for α and κ can be obtained using the large sample normality and independence of α and κ . For n large, $\sqrt{n}(\hat{\kappa}-\kappa) \sim N(0,1/B'(\kappa))$ and, when $\alpha > 0$, $\sqrt{n}(\hat{\alpha}-\alpha) \sim N(0,1/(\kappa B(\kappa)))$. So, estimating κ with $\hat{\kappa}$, an approximately 95% confidence interval for α is $[6.1^\circ, 15.3^\circ]$ and for κ is $[12.7, 25.9]$.

The estimates for the modal directions based on the MLE's are $(39^\circ, 135.7^\circ)$ and $(39.5^\circ, 170^\circ)$. The non-parametric tests, described in 3.4(i), were performed using these estimates for the modal directions, to assess the appropriateness of the double symmetry assumption. It was found that $r = 14$ and $w.\sqrt{3}/(19\times14\times34) = 0.0364$. Under the hypothesis of double symmetry, these statistics have distributions $Bi(\frac{1}{2}, 33)$ and $N(0.1)$ respectively. So the values of r and w are quite compatible with the double symmetry assumption.

These estimates of the modal directions are in fairly reasonable agreement with results obtained from a more recent investigation by Schmidt and McDougall (1977) (they give $(42.3^\circ, 123.5^\circ)$ and $(39.3^\circ, 174.5^\circ)$), except that in the later investigation the angle between the modal directions is somewhat larger. λ_1, the test statistic for $H_1 : v_1 = (42.3^\circ, 123.5^\circ)$ and $v_2 = (39.3^\circ, 174.5^\circ)$ versus $G : D$, unrestricted, was calculated as 8.8. Hence the null hypothesis, H_1, is (just) accepted at the 5 per cent level, and the two investigations appear to corroborate each other at least to some extent.

An analysis was also performed using the Fisher-Bingham family. Maximum likelihood estimates were obtained under the FDW_6 and FDW_5 hypotheses. However, in both cases the distributions fit were unimodal (for FDW_6, $\hat{\tau} = 0.97$ and for FDW_5, $\hat{\tau} = 0.99$; see (3.5.2) for the definition of τ) though they were close to the 'borderline' between unimodal and bimodal distributions. We should bear in mind that, since FDW_5 and FDW_6 contain unimodal distributions which do not have rotational symmetry, there will be a tendency for 'best-fit' distributions to be unimodal when the angle between the modes is small. So the values of $\hat{\tau}$ are not that surprising, since the angle between the modes, as estimated using D, is in fact quite small.

We conclude by suggesting that analyses using the D family are preferable (and far simpler) in cases in which there is a priori, or graphical, support for an assumption of bimodality, except when there is a considerable difference in the mode magnitudes, in which case FDW_6 should be used.

CHAPTER 4

THE SIMULATION OF DISTRIBUTIONS IN THE FISHER-BINGHAM FAMILY

4.1 Introduction

We return now to the Fisher-Bingham family, described in Chapter 2, and discuss how one might draw psuedo random samples from these distributions. As, in the most general case, there are five shape parameters involved, and there is no obvious way to proceed, it appears that compromises of one sort or another have to be made. Two distinct approaches, vaguely specified, are as follows: one might look for a computationally quick-and-simple acceptance-rejection procedure, likely to be substantially more efficient than those based on the crudest envelopes, but which perhaps, in the more extreme cases, would have small acceptance ratio. Alternatively, one might look for a globally efficient procedure, in which a substantial amount of prior computation may be required.

As long as the acceptance ratio is not too small, the first approach will be preferable if a small-to-moderate simulation study is to be performed, and the second will be preferable in sufficiently large simulation studies.

Most of our attention is focussed on the first approach.

A relatively simple two-stage simulation procedure is presented. The first stage involves simulating from the marginal distribution of the colalitude, θ , in an appropriately chosen coordinate system, and the second stage involves simulating from the conditional distribution of ϕ (the longitude) given θ . Envelopes for the marginal densities of $u = \cos\theta$, and ϕ given u (which, in the most general case, has a circular Fisher-Bingham distribution), are presented in section 4.2. The envelopes for the marginal density of u are based on the marginal Fisher distribution, which has density

$$f^*(u|\kappa^*,\alpha^*) = \left[\kappa^*/(2\sinh\kappa^*)\right] I_o(\kappa^*\sin\alpha^* v)\exp\{\kappa^*\cos\alpha^* u\} \tag{4.1.1}$$

where I_o is the modified Bessel function of degree zero, κ^* and α^* are the parameters, $v = (1-u^2)^{\frac{1}{2}}$, and the differential is du . If we were to put $\theta = \cos^{-1}u$, then (4.1.1) would be the marginal density of the colatitude of a unit vector from a Fisher distribution with concentration parameter κ^* , and mean direction with colatitude α^* . The envelopes for the circular Fisher-Bingham distribution are based on the von Mises distribution.

As the technical details will not be spelt out in later sections, we mention here that it is easy to generate a variable u with marginal Fisher distribution (4.1.1): it simply involves generating a Fisher unit vector, with concentration parameter κ^* , by inversion (Mardia (1972,

p. 232)) and then calculating the scalar product of the vector so generated with $(\sin\alpha^*, 0, \cos\alpha^*)'$. The generation of von Mises variables is less straightforward, but globally efficient methods are available (Best and Fisher (1979), Ulrich (1984)).

In section 4.3 the procedure is described and in section 4.4 special cases of interest are discussed. Indications are given as to what happens to the acceptance ratio in limiting cases. As might be expected, the acceptance ratio is not bounded away from zero.

In section 4.5, it is noted that any Fisher-Bingham distribution can be represented as a mixture of Fisher distributions. Implications of this observation are mentioned briefly. However, the mixing distribution is fairly complicated, so the mixture representation does not appear to be of much use from the point of view of simulation. Finally, in section 4.6, a summary and a further discussion are given.

Note: in an earlier version of this Chapter, and in Wood (1984), envelopes based on the bimodal distribution described in Chapter 3 were suggested. The referee for Wood (1984) made some helpful general suggestions which have led to a substantial change in the method proposed. I am responsible for all the technicalities in the new version, such as the formulation and proof of the Lemmas in section 4.2. Nevertheless, I am pleased to acknowledge his contribution.

4.2 Some Envelopes

In this section we shall obtain envelopes for an appropriate marginal density of a general Fisher-Bingham distribution. Envelopes will also be given for the 2-dimensional analogue of the Fisher-Bingham distribution. As a preliminary, we shall introduce several inequalities which are needed in the derivations of the envelopes. It is likely that these inequalities are well-known, but we have been unable to find suitable references, so their validity will also be proved.

Lemma (4.2.1) If $\rho \geq 0$, then

$$(e^{\rho u} + e^{-\rho u})/(1 + e^{-2\rho}) \geq e^{\rho u^2} \quad \text{for} \quad u \in [-1,1] .$$

Proof Because of the symmetry, we need only consider $u \in [0,1]$. If $\rho = 0$, there is nothing to prove, so assume ρ is positive and fixed. Consider the function $g(u) = \log(e^{\rho u} + e^{-\rho u}) - \rho u^2$ and its derivatives $g'(u) = \rho\sinh(\rho u)/\cosh(\rho u) - 2\rho u$ and $g''(u) = \rho^2/\cosh^2(\rho u) - 2\rho$. Two cases can arise.

(i) $\rho \leq 2$.

Here, $g''(u) < 0$ for $u \in (0,1]$. So $g'(u)$ is strictly decreasing for $u \in (0,1]$ and therefore has only one zero in $[0,1]$, at $u = 0$; this zero corresponds to a stationary maximum of g . Therefore $g(u)$ takes its minimum value on $[0,1]$ at $u = 1$.

(ii) $\rho > 2$.

It can be checked that there is a $\psi \in (0,1)$ such that $g''(u) > 0$ when $u \in (0,\psi)$ and $g''(u) < 0$ when $u \in (\psi,1]$. Therefore g can not have a minimum on $(0,1)$ since $g'(u)$ is strictly positive on $(0,\psi]$. It happens that g has a stationary minimum at $u = 0$; but it is easy to see that $g(1) \leq g(0)$.

So in both (i) and (ii) the minimum occurs at $u = 1$, and the result follows. 000

<u>Lemma (4.2.2)</u> If $\rho \geq 0$, then

$$2\left\{(1+e^{-2\rho})I_o(\rho/2)\right\}^{-1}e^{\rho/2}I_o(\rho v) \geq e^{\rho v^2} \quad \text{for } v \in [-1,1] .$$

<u>Proof</u> Let $u = v\cos\phi$ in (4.2.1). Then after putting $\cos^2\phi = (1+\cos 2\phi)/2$ on the RHS of (4.2.1), and integrating both sides with respect to ϕ over $[0,\pi]$, it is seen that

$$I_o(\rho v)/(1+e^{-2\rho}) \geq e^{\rho v^2/2}I_o(\rho v^2/2)/2 \quad . \tag{4.2.3}$$

Since the function $e^{-x}I_o(x)$ is monotonic decreasing for $x \geq 0$, it follows that

$$I_o(\rho v^2/2) \geq e^{\rho v^2/2}e^{-\rho/2}I_o(\rho/2) \quad , \tag{4.2.4}$$

and (4.2.3) and (4.2.4) together yield (4.2.2). 000

Lemma (4.2.5)

(i) For $x,y \geq 0$, $I_o(x+y) \geq I_o(x)I_o(y)$

(ii) $\{I_o(\rho)\}^{-1}[I_o(2\rho)]^{\frac{1}{2}}I_o(\rho v) \geq [I_o(2\rho v)]^{\frac{1}{2}}$ for $v \in [-1,1]$.

Proof of (i) This follows easily from the monotonicity of the functions I_o and $A(y) = \partial(\log I_o(y))/\partial y$.

Proof of (ii) When $\rho = 0$, there is nothing to prove, since $I_o(0) = 1$; and because I_o is an even function, we can restrict attention to $v \in [0,1]$. For fixed $\rho > 0$, consider the function $h(v) = 2\log(I_o(\rho v)) - \log(I_o(2\rho v))$. Since for $v \in (0,1]$, $h'(v) = 2[A(\rho v) - A(2\rho v)] < 0$, h takes its minimum value in $[0,1]$ at $v = 1$. Therefore

$$I_o(2\rho)[I_o(\rho v)/I_o(\rho)]^2 \geq I_o(2\rho v)$$

and (4.2.5) part (ii) follows when square roots are taken.

000

We shall now obtain some envelopes. Consider the Fisher-Bingham distribution for convenience parametrised as follows:

$$f(u,\phi|\lambda) = [c(\lambda)]^{-1}\exp\{\lambda_1 v\cos\phi + \lambda_2 v\sin\phi + \lambda_3 u + \lambda_4 v^2\cos 2\phi + \lambda_5 u^2\} . \quad (4.2.6)$$

The ranges of u and ϕ are $[-1,1]$ and $[0,2\pi)$ respectively; $v = (1-u^2)^{\frac{1}{2}}$, $\lambda = (\lambda_1,\ldots\lambda_5)$ is the vector of parameters, and, without loss of generality,

λ_1, λ_2, λ_3 and λ_4 are assumed to be non-negative; and the differential is $dud\phi$. Usually one would use the colatitude, $\theta = \cos^{-1}u$, instead of u.

Let $g(u)$ be the marginal distribution of u, i.e.

$$g(u) = \int_{\phi=0}^{2\pi} f(u,\phi|\lambda)d\phi \quad , \qquad u \varepsilon [-1,1] .$$

Two general envelopes will be obtained for $g(u)$, one for when λ_5 is positive, and the other for when it is negative. Putting $f_1(u,\phi) = \exp\{\lambda_1 v\cos\phi + \lambda_2 v\sin\phi\}$ and $f_2(u,\phi) = \exp\{\lambda_4 v^2\cos 2\phi\}$ and using the Cauchy-Schwartz inequality, it is seen that

$$(2\pi)^{-1}\int_{\phi=0}^{2\pi} f_1 f_2 d\phi \leqslant \left\{ I_o\left[2v(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}\right] I_o(2\lambda_4 v^2)\right\}^{\frac{1}{2}} . \qquad (4.2.7)$$

Since I_o is a monotonic increasing function and $v \varepsilon [0,1]$, we can replace $I_o(2\lambda_4 v^2)$ by $I_o(2\lambda_4 v)$ and still preserve the inequality. So using part (ii) of Lemma (4.2.5) on $\left\{I_o\left[2v(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}\right]\right\}^{\frac{1}{2}}$ and $\left\{I_o(2\lambda_4 v)\right\}^{\frac{1}{2}}$, and then using part (i) of Lemma (4.2.5), a more convenient bound for the LHS of (4.2.7) is obtained, namely

$$B(\lambda_1,\lambda_2,\lambda_4) I_o\left\{\left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} + \lambda_4\right]v\right\}$$

where $B(\lambda_1,\lambda_2,\lambda_4) =$

$$2\pi\left[I_o(2(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}})I_o(2\lambda_4)\right]^{\frac{1}{2}}/\{I_o((\lambda_1^2+\lambda_2^2)^{\frac{1}{2}})I_o(\lambda_4)\} \qquad (4.2.8)$$

Note: if λ_4 or $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}$ is zero, take $B = 2\pi$.

Two cases now arise, $\lambda_5 \geqslant 0$ and $\lambda_5 < 0$.

(i) $\lambda_5 \geqslant 0$. After using Lemma (4.2.1) on $\exp(\lambda_5 u^2)$, the following inequality is obtained:

$$g(u) \leq \{c(\lambda)\}^{-1} B(\lambda_1, \lambda_2, \lambda_4)\{1+\exp(-2\lambda_5)\}^{-1} \times \ldots$$

$$\ldots \times I_o([(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}+\lambda_4]v)[\exp\{(\lambda_3+\lambda_5)u\}+ \ldots$$

$$\ldots + \exp\{(\lambda_3-\lambda_5)u\}] \quad . \qquad (4.2.9)$$

The RHS of (4.2.9) is proportional to a discrete mixture of two marginal Fisher distributions, with the modal directions of the two original Fisher distributions lying on the same axis.

(ii) $\lambda_5 < 0$. From Lemma (4.2.2), it is seen that

$$D(\lambda_5) I_o(-\lambda_5 v) \geq \exp(\lambda_5 u^2) \qquad (4.2.10)$$

where $D(\lambda_5) = 2\exp(\lambda_5/2)\left\{(1+\exp(2\lambda_5))I_o(\lambda_5/2)\right\}^{-1}$. So, in this case, an envelope for $g(u)$ is given by

$$\{c(\lambda)\}^{-1} B(\lambda_1, \lambda_2, \lambda_4) D(\lambda_5) \times \ldots$$

$$\ldots \times I_o\{[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}+\lambda_4-\lambda_5]v\}\exp(\lambda_3 u) \qquad (4.2.11)$$

which is of marginal Fisher form.

We shall discuss the use of envelopes (4.2.9) and (4.2.11) in the general case in section 4.3, and specialise to Fisher-Bingham subfamilies of interest in section 4.4.

To complete the simulation procedure, it will be necessary to have a method for generating ϕ , conditional on the value of a previously generated u . The conditional distribution of ϕ can be obtained from (4.2.6). Four cases arise.

(a) $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} = \lambda_4 = 0$.

In this case, ϕ is uniform on $[0,2\pi)$.

(b) $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} \neq 0$, $\lambda_4 = 0$.

Here, ϕ given u is von Mises.

(c) $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} = 0$, $\lambda_4 \neq 0$.

In this case, 2ϕ given u is von Mises.

(d) $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} \neq 0$, $\lambda_4 \neq 0$.

In this case, ϕ given u has a 2-dimensional Fisher-Bingham distribution.

Case (a) poses no problem; cases (b) and (c) can be dealt with using the procedure described in Best and Fisher (1979), in which a wrapped Cauchy envelope is used. An alternative would be to use the procedure described in Ulrich (1984).

As case (d), the simulation of the 2-dimensional Fisher-Bingham distribution, appears not to have been discussed in the literature, we shall suggest a procedure here. The acceptance ratio for the procedure will not be globally bounded away from zero. However, it should be reasonably efficient for all but the very highly concentrated distributions in the family.

Consider the general 2-dimensional Fisher-Bingham density

$$h(\phi\,|\,\alpha,\beta_1,\beta_2) = \left\{d(\alpha,\beta_1,\beta_2)\right\}^{-1}\exp\left\{\beta_1\cos\phi+\beta_2\cos\left[2(\phi-\alpha)\right]\right\} \tag{4.2.12}$$

where $\phi \in [0,2\pi)$, $\beta_1,\beta_2 \geq 0$, and $\alpha \in [0,\pi/2]$. Three envelopes will be presented.

(i) (for $\alpha \geq \pi/2$)

$$h(\phi) \leq \left\{d(\alpha,\beta_1,\beta_2)\right\}^{-1}\exp\{\beta_2\}\exp\{\beta_1\cos\phi\} \tag{4.2.13}$$

(ii) (for $\alpha < \pi/2$)

With an application of Lemma (4.2.1) it can be seen that

$$\begin{aligned} h(\phi) \leq {} & \left\{d(\alpha,\beta_1,\beta_2)\right\}^{-1}\exp\left\{\beta_2\sin2\alpha+(\beta_2\cos2\alpha)/2\right\} \times \ldots \\ & \ldots \times \left\{1+\exp(-\beta_2\cos2\alpha)\right\}^{-1}\left[\exp\left\{(\beta_1+(\beta_2\cos2\alpha)/2)\cos\phi\right\} + \ldots \right. \\ & \left. \ldots + \exp\left\{(\beta_1-(\beta_2\cos2\alpha)/2)\cos\phi\right\}\right] \end{aligned} \tag{4.2.14}$$

(iii) Transform to $\psi = \phi - \alpha$. Then, using Lemma (4.2.1) again, it can be seen that

$$\begin{aligned} h(\psi+\alpha) \leq {} & \left\{d(\alpha,\beta_1,\beta_2)\right\}^{-1}\exp\left\{\beta_2/2\right\}\left\{1+\exp(-\beta_2)\right\}^{-1} \times \ldots \\ & \ldots \times \left[\exp\left\{(\beta_1\cos\alpha+\beta_2/2)\cos\psi-\beta_1\sin\alpha\sin\psi\right\} + \ldots \right. \\ & \left. \ldots + \exp\left\{(\beta_1\cos\alpha-\beta_2/2)\cos\psi-\beta_1\sin\alpha\sin\psi\right\}\right] . \end{aligned} \tag{4.2.15}$$

Envelope (i) is proportional to a von Mises density, and envelopes (ii) and (iii) are each proportional to mixtures of two von Mises densities. We recommend (iii) as an all-purpose envelope, and, in particular, it should usually perform better on average than envelopes (i) and (ii) when the parameters α , β_1 and β_2 in (4.2.12) are different for each ϕ generated. However, there do exist particular cases in which (i) and (ii) have larger acceptance ratio than (iii).

4.3 The Simulation Procedure

We begin this section with a general point. Suppose that we transform in (4.2.6) to cartesian cooordinate,

$$x = v\cos\phi \ , \quad y = v\sin\phi \quad \text{and} \quad z = u \ .$$

Then it is seen that the eigenvectors of the 'Bingham' matrix (that is, the eigenvalues of the matrix A in (2.1.1)) are parallel to the coordinate axes. Since λ_4 in (4.2.6) is assumed non-negative, there are three (as opposed to six) ways of reducing a general Fisher-Bingham density to the form (4.2.6) by orthogonal transformation, each one corresponding to the alignment of a 'Bingham' eigenvector with the Z-axis. In each case, the X-axis lies parallel to the remaining 'Bingham' eigenvector with larger eigenvalue, so that $\lambda_4 \geq 0$ is satisfied. The positive directions on the coordinate axes are also determined since $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$.

In general, these three permutations of the coordinate axes lead to different envelopes and different acceptance ratios. Therefore, one should check which of the three has largest acceptance ratio, and then choose the corresponding envelope. In what follows, we shall assume implicitly that this has been done.

The simulation procedure will now be described. The general case will be dealt with first, and then in section 4.4 the procedure will be discussed further in four special cases of interest: those in which the distribution to be simulated lies in either FB_4, FB_5, FDW_6 or the Bingham subfamily. Throughout, it will be assumed that the Fisher-Bingham density to be simulated is of the form (4.2.6) with parameters $\lambda_1, \ldots \lambda_5$; and that $g(u)$ is the marginal density of u.

(i) The General Case $(\lambda_5 > 0)$

From (4.2.9), it is seen that the envelope for $g(u)$ is proportional to a mixture of two marginal Fisher distributions. This envelope is given by

$$(a^*)^{-1}\left[p^* f^*(u|\kappa_1^*, \alpha_1^*) + (1-p^*) f^*(u|\kappa_2^*, \alpha_2^*)\right] \tag{4.3.1}$$

where f^* is the marginal Fisher density given in (4.1.1), p^* is the mixing proportion and a^* is the acceptance ratio. The envelope parameters κ_1^*, α_1^*, κ_2^*, α_2^*, p^* and a^* are given by:

$$\kappa_1^* = \left\{ 2\left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}\lambda_4 + \lambda_3\lambda_5\right] + \sum_{j=1}^{5} \lambda_j^2 \right\}^{\frac{1}{2}} \tag{4.3.2}$$

$$\alpha_1^* = \tan^{-1}\left\{ \left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} + \lambda_4\right] / \left[\lambda_3 + \lambda_5\right] \right\} \tag{4.3.3}$$

$$\kappa_2^* = \left\{ 2\left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}\lambda_4 - \lambda_3\lambda_5\right] + \sum_{j=1}^{5} \lambda_j^2 \right\}^{\frac{1}{2}} \tag{4.3.4}$$

$$\alpha_2^* = \tan^{-1}\left\{\left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}+\lambda_4\right] / \left[\lambda_3-\lambda_5\right]\right\} \tag{4.3.5}$$

$$p^* = [\kappa_2^*\sinh\kappa_1^*]/[\kappa_1^*\sinh\kappa_2^* + \kappa_2^*\sinh\kappa_1^*] \tag{4.3.6}$$

$$a^* = p^*\kappa_1^* c(\lambda)[1+\exp(-2\lambda_5)]/[2B(\lambda_1,\lambda_2,\lambda_4)\sinh\kappa_1^*] \tag{4.3.7}$$

In (4.3.7), $c(\lambda)$ is the normalising constant in (4.2.6) and the function $B(\lambda_1,\lambda_2,\lambda_4)$ is defined in (4.2.8).

The simulation procedure is as follows:

<u>Procedure (i)</u>

Step 0. Calculate $\kappa_1^*, \alpha_1^*, \kappa_2^*, \alpha_2^*, p^*$ and a^* using (4.3.2)-(4.3.7).

Step 1. Generate a uniform variable $s \in [0,1]$. If $s \leq p^*$ put $\kappa^* = \kappa_1^*$, $\alpha^* = \alpha_1^*$; if $s > p^*$, put $\kappa^* = \kappa_2^*$, $\alpha^* = \alpha_2^*$.

Step 2. Generate a variable $u \in [-1,1]$ from a marginal Fisher distribution with parameters κ^* and α^* (calculated in Step 1).

Step 3. Generate a uniform variable $t \in [0,1]$ and check whether

$$a^* g(u) / \left[p^* f^*(u \mid \kappa_1^*, \alpha_1^*) + (1-p^*) f^*(u \mid \kappa_2^*, \alpha_2^*)\right] \geq t \tag{4.3.8}$$

is satisfied. If it is, accept u ; otherwise, return to Step 1. (Note: the calculation of g and f^* will be discussed at the end of the section.)

(ii) The General Case ($\lambda_5 \leq 0$)

From (4.2.11), it is seen that the envelope for $g(u)$ is proportional to a marginal Fisher distribution. This envelope is given by

$$(a^*)^{-1} f^*(u|\kappa^*, \alpha^*) \tag{4.3.9}$$

where, as before, f^* is the marginal Fisher density and a^* is the acceptance ratio. The envelope parameters κ^*, α^* and a^* are given by:

$$\kappa^* = \left\{2\left[(\lambda_4-\lambda_5)(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}-\lambda_4\lambda_5\right]+\sum_{j=1}^{5} \lambda_j^2\right\}^{\frac{1}{2}} \tag{4.3.10}$$

$$\alpha^* = \tan^{-1}\left\{\left[(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}+\lambda_4-\lambda_5\right]/\lambda_3\right\} \tag{4.3.11}$$

$$a^* = \kappa^* c(\lambda)/\left[2B(\lambda_1,\lambda_2,\lambda_4)D(\lambda_5)\sinh\kappa^*\right] \tag{4.3.12}$$

where again $c(\lambda)$ is the normalising constant in (4.2.6), and B and D are the functions defined (4.2.8) and (4.2.10) respectively. In this case, the simulation procedure is:

Procedure (ii)

Step 0. Calculate κ^*, α^* and a^* using (4.3.10)-(4.3.12).

Step 1. Generate a variable $u \varepsilon [-1,1]$ from a marginal Fisher distribution with parameters κ^* and α^* (calculated in Step 0).

Step 2. Generate a uniform variable $s \varepsilon [0,1]$ and check whether

$$a^* g(u)/f^*(u|\kappa^*,\alpha^*) \geq s \qquad (4.3.13)$$

is satisfied. If it is, accept u ; otherwise return to Step 1.

We now specify how to generate the longitude, ϕ , for a given value of u . The von Mises density will be denoted by

$$m^*(\phi|\sigma^*,\gamma^*) = \left\{2\pi I_0(\sigma^*)\right\}^{-1} \exp\left\{\sigma^* \cos(\phi-\gamma^*)\right\} \qquad (4.3.14)$$

where $\phi \varepsilon [0,2\pi)$. When $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}} = 0$, the distribution of 2ϕ given u is von Mises; and when $\lambda_4 = 0$, the distribution of ϕ given u is von Mises. In both cases it is easy to obtain the appropriate values of σ^* and γ^* in terms of $\lambda_1, \lambda_2, \lambda_4$ and the given value of u . For convenience we shall omit these formulae, and just deal with the case in which both $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}$ and λ_4 are non-zero. When this is the case, we recommend that envelope (4.2.15) be used unless there is concrete evidence to suggest that envelope (4.2.13) or (4.2.14) will perform better.

For fixed u , define

$$v = (1-u^2)^{\frac{1}{2}} \; ; \; \lambda_1^* = \lambda_1 v \; , \; \lambda_2^* = \lambda_2 v \; , \; \lambda_4^* = \lambda_4 v^2 \; ;$$

$$\eta^* = \left[(\lambda_1^*)^2+(\lambda_2^*)^2\right]^{\frac{1}{2}} \; , \; \delta^* = \tan^{-1}(\lambda_2^*/\lambda_1^*) \; . \tag{4.3.15}$$

Then the conditional density of ϕ given u is, in the notation of (4.2.12), $h(\phi+\delta^*|-\delta^*,\eta^*,\lambda_4^*)$. The envelope (based on (4.2.15)) for this density is:

$$(r^*)^{-1}\left[q^* m^*(\phi\,|\,\sigma_1^*,\gamma_1^*)+(1-q^*)m^*(\phi\,|\,\sigma_2^*,\gamma_2^*)\right] \tag{4.3.16}$$

where m^* is the von Mises density given in (4.3.14), r^* is the acceptance ratio for given u and q^* is the mixing proportion. The envelope parameters σ_1^*, γ_1^*, σ_2^*, γ_2, q^* and r^* are given by:

$$\sigma_1^* = (\eta^*)^2+(\lambda_4^*)^2+\eta^*\lambda_4^*\cos\delta^* \tag{4.3.17}$$

$$\gamma_1^* = \tan^{-1}\{[\eta^*\cos\delta^*+\lambda_4^*/2]/[-\eta^*\sin\delta^*]\} \tag{4.3.18}$$

$$\sigma_2^* = (\eta^*)^2+(\lambda_4^*)^2-\eta^*\lambda_4^*\cos\delta^* \tag{4.3.19}$$

$$\gamma_2^* = \tan^{-1}\{[\eta^*\cos\delta^*-\lambda_4^*/2]/[-\eta^*\sin\delta^*]\} \tag{4.3.20}$$

$$q^* = I_o(\sigma_1^*)/[I_o(\sigma_1^*)+I_o(\sigma_2^*)] \tag{4.3.21}$$

$$r^* = q^* d(-\delta^*,\eta^*,\lambda_4^*)\exp(-\lambda_4^*/2)[1+\exp(-\lambda_4^*)] \tag{4.3.22}$$

In (4.3.17) - (4.3.22), λ_4^*, η^* and δ^* are defined in (4.3.15); in (4.3.21), I_o is the modified Bessel function of degree zero; and in (4.3.22), d is the normalising constant in (4.2.12). The procedure for generating a pair (u,ϕ) from the Fisher-Bingham distribution (4.2.6) can now be given.

Procedure (iii)

Step 0. Generate $u \in [-1,1]$ using Procedure (i) if $\lambda_5 > 0$. If $\lambda_5 \leq 0$, use Procedure (ii).

Step 1. Calculate $\lambda_4^*, \eta^*, \delta^*, \sigma_1^*, \gamma_1^*, \sigma_2^*, \gamma_2^*, q^*$ and r^* using (4.3.15) and (4.3.17)-(4.3.22).

Step 2. Generate a uniform variable $s \in [0,1]$. If $s \leq q^*$, put $\sigma^* = \sigma_1^*$, $\gamma^* = \gamma_1^*$. If $s > q^*$, put $\sigma^* = \sigma_2^*$, $\gamma^* = \gamma_2^*$.

Step 3. Generate a von Mises variable $\phi \in [0,2\pi)$ with distribution parameters σ^* and γ^* (calculated in Step 2).

Step 4. Generate a uniform variable $t \in [0,1]$. Check whether

$$r^* h(\phi+\delta^* | -\delta^*, \eta^*, \lambda_4^*)/[q^* m^*(\phi|\sigma_1^*, \gamma_1^*) + (1-q^*) m^*(\phi|\sigma_2^*, \gamma_2^*)] \quad (4.3.23)$$

is satisfied. If it is, accept ϕ. Otherwise, return to Step 2.

Then (u,ϕ) will have been generated from the Fisher-Bingham distribution (4.2.6).

Comment: in evaluating the LHS of (4.3.23), there is no need to calculate the normalising constant $d(-\delta^*, \eta^*, \lambda_4^*)$ since it cancels out. This comment also applies to $c(\lambda)$ in (4.3.8) and (4.3.13).

We return now to the evaluation of the LHS of (4.3.8). Note: all that we say below also applies to the LHS of (4.3.13); but for simplicity we assume $\lambda_5 > 0$. A speedy method of evaluating the modified Bessel function I_0 is required, since this is needed in the evaluation of f^* (the marginal Fisher density), and g (the marginal distribution of u) if either $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}$ or λ_4 is zero. We have used NAG subroutine S18AEF. If neither $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}$ or λ_4 is zero, the evaluation of g is more problematic, because it involves the evaluation of d, the normalising constant for the circular Fisher-Bingham distribution. However, there is a short cut available. To see this, observe that

$$g(u) = \{c(\lambda)\}^{-1}\exp\{\lambda_3 u+\lambda_5 u^2-\lambda_4 v^2\}w(u;\lambda_1,\lambda_2,\lambda_4)$$

where $w(u;\lambda_1,\lambda_2,\lambda_4) = \int_0^{2\pi} \exp\{\lambda_1 v\cos\phi+\lambda_2 v\sin\phi+2\lambda_4 v^2\cos^2\phi\}d\phi$.
It is not difficult to check that, when λ_1, λ_2 and λ_4 are non-negative, $w(u;\lambda_1,\lambda_2,\lambda_4)$ is an even function of u and a monotonic decreasing function of $|u|$.

We suggest the following: as a preliminary, calculate $w(\xi_j;\lambda_1,\lambda_2,\lambda_4)$ at appropriate points $0 = \xi_o < \xi_1 \ldots < \xi_N < \xi_{N+1} = 1$. It is desirable that $\xi_1,\ldots\xi_N$ be chosen such that

$$\mathrm{Prob}(|u| \geq \xi_j) \simeq j/(N+1) \quad .$$

Then for a given u (generated in Step 1 of Procedure (i)) find the integer p such that $\xi_{p-1} < |u| \leq \xi_p$, and then replace $w(u;\lambda_1,\lambda_2,\lambda_4)$ in $g(u)$ by $w(\xi_p;\lambda_1,\lambda_2,\lambda_4)$. If (4.3.8) is satisfied, then accept u ; otherwise, replace $w(u;\lambda_1,\lambda_2,\lambda_4)$ in $g(u)$ by $w(\xi_{p-1};\lambda_1,\lambda_2,\lambda_4)$. If (4.3.8) is still not satisfied, reject u ; otherwise evaluate $w(u;\lambda_1,\lambda_2,\lambda_4)$ using the formula presented in comment (i) of section 2.2. This involves the computation of Bessel function ratios, and can be done comparatively quickly using the method proposed in Amos (1974).

In effect, the test of whether (4.3.8) holds has been decomposed into a quick check and a slow check, such that the quick check will be used with high probability, and the slow check with low probability. When $\xi_1,\ldots\xi_N$ are equally spaced on the probability scale, the probability that the slow check will be used is N^{-1} . The larger the choice for N , the more initial computation will be required (though note that $\xi_1,\ldots\xi_N$ will only need to be computed once for given λ_1, λ_2 and λ_4); but the greater the probability that the quick check will be used. For many purposes, $N = 64$ or $N = 128$ should be a reasonable choice for N .

4.4 The Procedure in Some Special Cases of Interest

The simulation procedure will now be discussed in more detail in four special cases of interest: those in which the Fisher-Bingham distribution to be simulated lies in either FB_4 , FB_5 , FDW_6 or the Bingham subfamily. In particular, the behaviour of the acceptance ratio in the simulation of the marginal distribution of u is indicated. The Dimroth-Watson subfamily is the only one in which the acceptance ratio is bounded away from zero. In each of the other cases, the acceptance ratio will in general converge to zero when the shape of the distribution to be simulated is held fixed, but the concentration about the mode(s) is allowed to tend to infinity. It is not difficult to see why this should happen: in general, the mode(s) of the envelope will not exactly coincide with those of the distribution to be simulated, and for sufficiently highly concentrated distributions one pays the price. We know of no way around this except at a heavy cost in prior computation (see section 4.6).

As above, $\lambda_1, \ldots \lambda_5$ refer to the parameters in (4.2.6). In none of the subfamilies discussed below are both $(\lambda_1^2+\lambda_2^2)^{\frac{1}{2}}$ and λ_4 non-zero; therefore we can take $B(\lambda_1, \lambda_2, \lambda_4)$ to be 2π .

(a) FB_4 $(\lambda_1 = \lambda_2 = \lambda_4)$

When $\lambda_5 < 0$, u has a Normal distribution, with mean $\lambda_3/[2(-\lambda_5)^{\frac{1}{2}}]$ and variance $-1/\lambda_5$, which is truncated at

$u = \pm 1$. So when $\lambda_5 < 0$, an alternative simulation procedure is to generate u from the Normal distribution with mean and variance given above, and accept it if it lies in $[-1,1]$, but reject it otherwise. From (4.3.10) and (4.3.12), it can be seen that the condition for this alternative procedure to have greater acceptance ratio than the procedure based on envelope (4.2.11) is:

$$(\lambda_3^2+\lambda_5^2)^{\frac{1}{2}}/\{2D(\lambda_5)\sinh[(\lambda_3^2+\lambda_5^2)^{\frac{1}{2}}]\} < 1 \qquad (4.4.1)$$

where D is defined in (4.2.10).
Condition (4.4.1) will be satisfied if either λ_3 or $-\lambda_5$ is sufficiently large.

Let $\lambda_3 = \rho$, $\lambda_5 = \rho t$. Then holding t fixed and allowing ρ to vary is equivalent to keeping the distribution shape fixed, and allowing the concentration to vary. When $t \geq 0$ and envelope (4.2.9) is used, it can be checked that as ρ tends to infinity, the acceptance ratio tends to $(1+t)/(1+2t)$, which is always greater than $1/2$ for $t \geq 0$. We omit the technical details, which are straightforward. Therefore, for $\lambda_5 \geq 0$, the acceptance ratio is bounded away from zero.

When $t < 0$, two procedures are available: the one mentioned above, and that based on envelope (4.2.11). If the procedure with larger acceptance ratio is chosen, two cases arise: when $-1/2 < t < 0$, the acceptance ratio is not bounded away from zero; but when $t < -1/2$, it is.

The calculation of these positive lower bounds appears not to be straightforward. When $-1/2 < t < 0$ and ρ tends to infinity, the acceptance ratio converges to zero at the same rate as $\Phi\left[\rho^{\frac{1}{2}}((-t)^{\frac{1}{2}}-(-t)^{-\frac{1}{2}}/2)\right]$, where Φ is the standard Normal cumulative distribution function.

It should be clear from this discussion that the acceptance ratio for the Dimroth-Watson subfamily is bounded away from zero.

(b) FB_5 $(\lambda_1 = \lambda_2 = \lambda_5 = 0)$

In this case, $D = 1$. Let $\lambda_3 = \rho$, $\lambda_4 = \rho t$. For fixed t , the acceptance ratio converges to zero as ρ tends to infinity, but the rate at which it does so depends on t . When $0 < t < 1/2$, the unimodal case, it is not too difficult to show that the rate of convergence to zero is given by

$$\left[(1+t^2)/(1-4t^2)\right]^{\frac{1}{2}}\exp\{\rho\left[1-(1+t^2)^{\frac{1}{2}}\right]\} \quad .$$

When $t \geqslant 1/2$, the calculation is more complicated.

(c) FDW_6 $(\lambda_4 = 0)$

As in the other cases the acceptance ratio is not bounded away from zero. Because of the lack of symmetry of these distributions, it appears to be difficult to calculate the rate at which this occurs. However, it is straightforward to perform these calculations for unimodal FDW_5 distributions, and the rates are similar in form to those for FB_5 .

(d) The Bingham Subfamily $(\lambda_1 = \lambda_2 = \lambda_3 = 0)$

The acceptance ratio is not bounded away from zero except in the Dimroth-Watson case $(\lambda_4 = 0)$. The rates are, again, similar in form to those for FB_5.

In some of these cases it should certainly be worth 'permuting the axes', as described at the beginning of section 4.3, in order to find a better envelope. One would particularly expect this to be so for Bingham and bimodal FB_5 and FDW_5 distributions.

4.5 A Mixture Representation for the Fisher-Bingham family

We now briefly describe a mixture representation for the Fisher-Bingham family. Only the 3-dimensional case will be referred to, but all the results in this section have analogues in other dimensions. Although this representation seems not to be of much use from the point of view of simulation, because the mixing distribution appears to be difficult to simulate, it does have some interesting implications.

Consider the Fisher-Bingham density

$$c^{-1}\exp\left\{\kappa(x'\mu)+\sum_{j=1}^{3}\lambda_j(x'\nu_j)^2\right\} \qquad x \in S_3 \qquad (4.5.1)$$

and without loss of generality assume that λ_1, λ_2 and λ_3 are positive; and let $h(x,y)$ be the density on

$S_3 \times R^3$ given by

$$[(2\pi)^{3/2}c]^{-1}\exp\left\{\kappa(x'\mu)+\sum_{j=1}^{3}\lambda_j^{\frac{1}{2}}z_j(x'\nu_j)-\tfrac{1}{2}z'z\right\} \tag{4.5.2}$$

with $x \in S_3$ and $z = (z_1,z_2,z_3)' \in R^3$, and c the same as in (4.5.1). Then it is easy to check that

(a) $h(x|z)$, the density of x given z, is a Fisher density with parameters $\kappa(z) \in R$, $\mu(z) \in S_3$ given by

$$\kappa(z) = \left\{ (\kappa\mu + \sum_{j=1}^{3}\lambda_j^{\frac{1}{2}}z_j\nu_j)'(\kappa\mu + \sum_{j=1}^{3}\lambda_j^{\frac{1}{2}}z_j\nu_j)\right\}^{\frac{1}{2}} \tag{4.5.3}$$

and $\mu(z) = (\kappa\mu + \sum_{j=1}^{3}\lambda_j^{\frac{1}{2}}z_j\nu_j)/\kappa(z)$.

(b) The marginal density of x is precisely (4.5.1), a Fisher-Bingham density.

(c) $h(z|x)$, the density of z given x, is Normal.

(d) The marginal density of z is

$$[(2/\pi)^{\frac{1}{2}}\sinh(\kappa(z))/(c\kappa(z))]\exp\{-\tfrac{1}{2}z'z\} \tag{4.5.4}$$

with $\kappa(z)$ given in (4.5.3). So, the Fisher-Bingham density (4.5.1) is a mixture of Fisher distributions with parameters $\kappa(z)$ and $\mu(z)$, and the mixing distribution has density (4.5.4).

We now mention two implications of this representation of the Fisher-Bingham distribution. Firstly, by entirely elementary methods one can easily obtain a neat expression for the normalising constant of the Fisher-Bingham distribution.

Replacing $\sinh(\kappa(z))/\kappa(z)$ by its Taylor series

$$\sum_{r=0}^{\infty} \left[\kappa(z)\right]^{2r}/(2r+1)!$$

and integrating term by term with respect to z, it is seen that c, the normalising constant, can be expressed as a (weighted) sum of the moments of a non-central quadratic form in Normal variables. In fact, given a minor reparametrisation, this expression is identical to that given in De Waal (1979) and discussed in section 2.2 above, though he obtains it via some complicated distribution theory.

A second implication of the mixture representation is that any Fisher-Bingham distribution can be represented as a (fairly complicated) mixture of the spherical Brownian motion distributions, discussed in Roberts and Ursell (1960), though the mixing is both over 'fixed stopping time' and initial position. This follows immediately from results in Hartman and Watson (1974), who show that the Fisher distribution is a mixture over 'fixed stopping times' of spherical Brownian motions.

4.6 Summary and Discussion

In this Chapter a method for simulating distributions in the Fisher-Bingham family is presented. The required envelopes are derived in section 4.2 and details of the procedure are given in section 4.3. The proposed method is rather more simple and effective for the FB_4 , FB_5 , FDW_6 and Bingham subfamilies than it is in the general case. In none of these subfamilies is the acceptance ratio bounded away from zero, though the method should prove adequate if the distribution to be simulated is not too highly concentrated about its mode(s).

The question arises as to how, if one were prepared to perform heavy prior computation, one might arrive at a globally efficient procedure. One possibility would be to use the integrand in (2.2.3) as an alternative marginal distribution. The form of this distribution is too complicated for there to be much hope of deriving good envelopes by theoretical means. But it would, of course, be possible to obtain a good envelope numerically for any given parameter values, though presumably heavy initial computation would be required. However, since the conditional distribution of the longitude is always von Mises in this case, the second part of the simulation would be straightforward.

CHAPTER 5

ROBUST ESTIMATION OF THE CONCENTRATION PARAMETER OF A FISHER DISTRIBUTION

5.1 Introduction

It has been noted by several authors, including Mardia (1972), Watson (1973) and Collett (1978), that the maximum likelihood estimator of concentration, for both the von Mises distribution on the circle and the Fisher distribution on the sphere, can be dramatically influenced by extreme observations. This is especially true when the majority of observations are highly concentrated about a particular direction, which frequently happens with directional data occurring in practice. In many contexts, the concentration parameter may be viewed as a nuisance parameter; however, even in these situations it would still be desirable to have a robust estimator available, so as to avoid making inferences about the mean direction which are sensitive to changes in a small proportion of the observations.

A promising class of potentially robust estimators of concentration would seem to be the L-statistic estimators which were suggested by Fisher (1982) in a brief note. Here we follow Fisher's suggestion and investigate their properties in the spherical case in some detail.

An outline of the chapter is as follows. A brief survey of estimators of concentration, with reference to robustness properties, is given in section 5.2. In sections 5.3 and 5.4 it is shown that if the underlying distribution has oval symmetry, and mild regularity conditions are satisfied, then the asymptotic distributions of L-estimators of concentration do not depend on whether the mean direction is known or estimated. This result is not at all surprising, but its proof does require some work.

Results from a simulation study of the small sample properties of some L-estimators are presented in section 5.5. These indicate that the best of the L-estimators may be expected to perform rather well for a Fisher distribution with at least moderate concentration.

Finally, in section 5.6, the possibility of using estimators broadly of this type in other contexts is discussed, in particular as robust estimators of the eigenvalues of the covariance matrix of a multivariate Normal distribution.

5.2 A Brief Survey of Estimators of Concentration

We now briefly mention some estimators of the concentration parameter, κ , of a Fisher distribution, with reference to their robustness properties. To assess the robustness of estimators we have used their influence curves, and in particular the gross error sensitivity, which is defined to be the supremum of the absolute value of an influence curve (Hampel (1974)).

(i) The MLE and related estimators. Schou's estimator (Schou (1978)).

The maximum likelihood estimator (MLE) is given by

$$\hat{\kappa} = B^{-1}(R/n) \quad \text{where} \quad B(.) = \coth(.)-1/(.)$$

and R is the sample resultant length for n observations. The related estimators are of the form $\rho_n\hat{\kappa}$, where the multiplier ρ_n depends only on n and is introduced to reduce the bias of the MLE. Mardia (1972), McFadden (1980) and Best and Fisher (1981) have suggested

$$\rho_n = (n-1)/n \ , \ (n-2)/n \quad \text{and} \quad \{(n-1)/n\}^{3/2}$$

respectively. These estimators perform well when the Fisher model is correct, but their robustness properties leave something to be desired. In particular, the gross error sensitivity is approximately $2\kappa^2$ as κ increases. This supremum is, not surprisingly, attained at the point diametrically opposed to the mean direction.

The estimator suggested by Schou (1978) is the maximum likelihood estimator, $\tilde{\kappa}_S$ say, based on the marginal distribution of the resultant length, R, and is the solution of

$$B(\tilde{\kappa}_S) = (R/n)B(\tilde{\kappa}_S R) \quad .$$

This estimator has reasonably good small sample properties when the Fisher model is correct, and as $n \to \infty$ it becomes indistinguishable from the full MLE; however, it suffers from a similar lack of robustness.

(ii) <u>An eigenvalue estimator, $\tilde{\kappa}_E$</u>

This estimator is defined to be the solution of

$$\lambda/n = 1 - 2B(\tilde{\kappa}_E)/\tilde{\kappa}_E$$

where λ is the largest eigenvalue of the matrix of sums and products (see Mardia (1972, 8.4.16)). It is related, though not quite analogous, to Collett's (1978) estimator for the circular case.

$\tilde{\kappa}_E$ is certainly robust against extreme outliers, but the gross error sensitivity is again proportional to κ^2, when κ is not too small, the most influential directions being those perpendicular to the mean direction.

(iii) <u>Estimators based on the set of resultants of pairs of observations</u>

Suppose we are given an independent sample of unit vectors $x_1, \ldots x_n$ from a Fisher distribution. Then the set of resultants of pairs of observations is

$$\underline{R} = \{r_{ij} : 1 \leq i < j \leq n\} \quad \text{where} \quad r_{ij} = 2\cos(\theta_{ij}/2)$$

and $\theta_{ij} = \cos^{-1}(x_i{}'x_j)$. It so happens that the marginal distribution of r_{ij} is quite simple (Mardia (1972, 8.6.32)) though the joint distribution of the pairs (r_{ij}, r_{ik}) is rather complicated.

One plausible estimator of this type is based on the median, $\tilde{\psi}_n$ say, of the set $\underline{R}$. (This was suggested by T. Lewis in a private communication.) We would certainly expect this estimator to have good robustness properties. It can be shown that with probability 1

$$\tilde{\psi}_n \to \kappa^{-1}\cosh^{-1}([1+\cosh 2\kappa]/2) ,$$

the RHS being the median of the marginal distribution of r_{ij}; and $\tilde{\psi}_n$ is asymptotically Normal (using results in Serfling (1984)). However, the asymptotic variance of $\tilde{\psi}_n$ depends on the joint distribution of (r_{ij}, r_{ik}), and appears to be a very complicated function of κ. This seems to be a feature of estimators based on the set $\underline{R}$: their asymptotic variances are difficult to calculate.

We now discuss the L-estimators suggested by Fisher (1982). It will be helpful to introduce some notation. Let $\{F\}$ be the space of distribution functions on S_3, the unit sphere in 3 dimensions, and for any $F \in \{F\}$ and $\gamma \in S_3$, define

$$F_\gamma(t) = \text{prob}_F\{x \in S_3 : x'\gamma \geq t\} \quad \text{for} \quad t \in [-1,1] .$$

Consider the functionals $T : \{F\} \times S_3 \to R$ of the form

$$T(F,\gamma) = \int_{-1}^{1} tJ(F_\gamma(t))dF_\gamma(t) \tag{5.2.1}$$

for fixed $J : [0,1] \to R$. The choice of J will be discussed shortly.

Suppose we are given a sample $x_1,\ldots x_n$ of independent unit vectors from a Fisher distribution $F(\mu,\kappa)$. Let $\tilde{F}_n$ be the empirical distribution function for the sample, and $\tilde{\mu}_n$ some estimator of μ such as the sample mean or median direction. Then an L-estimator of κ is, by definition, of the form

$$\tilde{\kappa} = H^{-1}(T(\tilde{F}_n,\tilde{\mu}_n)) \quad \text{where} \quad H(\kappa) = T(F(\mu,\kappa),\mu). \qquad (5.2.2)$$

Now $T(\tilde{F}_n,\mu)$ is a linear combination of order statistics from a sample of independent, identically distributed variables. In sections 3 and 4 we show that

$$n^{\frac{1}{2}}\left[T(\tilde{F}_n,\tilde{\mu}_n) - T(\tilde{F}_n,\mu)\right] \to 0 \qquad (5.2.3)$$

in probability when the underlying distribution is Fisher, and mild restrictions are imposed on J . In fact, the same result also holds for any suitably well-behaved distribution with oval symmetry about μ . It is worth noting that if the underlying distribution does not have oval symmetry, then in general (5.2.3) is false.

An immediate consequence of (5.2.3) is that the asymptotic distribution of $n^{\frac{1}{2}}\tilde{\kappa}$, where $\tilde{\kappa}$ is defined in (5.2.2), does not depend on whether μ is known or estimated, so that in particular we can calculate the variance and influence curve of $\tilde{\kappa}$ as though μ were known.

We shall be considering three types of J-function, each parameterised by $a \in (0,1)$.

Case I (Trimmed estimators)

$J_1(y) = 0$ if $0 \leq y < a$ and $J_1(y) = (1-a)^{-1}$ if $a \leq y \leq 1$.

Case II (Quantile estimators)

$J_2(y) = \delta_a(y)$, where $\delta_a(y)$ is the Dirac delta function with unit mass at $y = a$.

Case III (Winsorised estimators)

$J_3(y) = (1-a)J_1(y) + aJ_2(y)$.

The corresponding functions H_1 , H_2 and H_3 , which will be monotonic because J_1 , J_2 and J_3 are non-negative, the functionals T_1 , T_2 and T_3 , and estimators $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ are defined by (5.2.2). Straightforward calculation yields the functions H_1 , H_2 and H_3 , and expressions for the asymptotic variances and influence curves for $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ can be obtained. In the situation of most practical significance, that in which κ , the concentration, is at least moderately large, e.g. $\kappa \geq 3$, it is reasonable to ignore terms of order $e^{-2\kappa}$, and in this case the formulae have simple approximations. However, when $\kappa < 3$, these approximations are likely to be unreliable. If there is some doubt as to their validity the function H should be inverted numerically to obtain the estimate. This will not be a problem for the three types of L-estimator we are considering, because H_1 , H_2 and H_3 have reasonably simple closed form.

The approximate formulae, for κ not too small, are given below. For convenience, we have presented the inverse functions $H_i^{-1}(.)$, rather than $H_i(.)$; and expressed the influence curves as functions of probability values, rather than quantiles on $[-1,1]$. $I(.)$ denotes the indicator function.

Case I

$$H_1^{-1}(y) = \{1+a\log a/(1-a)\}/(1-y)$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}T_1) = (2a\log a+1-a^2)/[\kappa^2(1-a)^2]$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}\tilde{\kappa}_1) = \kappa^2(2a\log a+1-a^2)/(1-a+a\log a)^2$$

$$IC_1(y) = \{\kappa/(1-a+a\log a)\}\{1-a+I(y\leq a)\log a + I(y>a)\log y\}$$

Case II

$$H_2^{-1}(y) = -\log a/(1-y)$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}T_2) = (1-a)/(\kappa^2 a)$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}\tilde{\kappa}_2) = \kappa^2(1-a)/[a(\log a)^2]$$

$$IC_2(y) = \{\kappa/(-a\log a)\}\{a-I(y\leq a)\}$$

Case III

$$H_3^{-1}(y) = (1-a)/(1-y)$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}T_3) = (1-a)/\kappa^2$$

$$\lim_{n\to\infty} \mathrm{var}(n^{\frac{1}{2}}\tilde{\kappa}_3) = \kappa^2/(1-a)$$

$$IC_3(y) = \{\kappa/(1-a)\}\{1-I(y\leq a) + I(y\leq a)\log a + I(y>a)\log y\}$$

The gross error sensitivity of each of $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and

$\tilde{\kappa}_3$ is proportional to κ. This is in contrast to estimators based on the sample resultant and the eigenvalue estimator, mentioned earlier, whose gross error sensitivities are proportional to κ^2. This indicates that these L-estimators will be substantially less influenced by observations lying some way from a concentrated cluster of points.

The maximum and minimum values of the influence curves, scaled by κ, and the asymptotic efficiencies of the three estimators relative to the maximum likelihood estimator, are given in Table 5.1. It can be seen that there is some, though not a dramatic, increase in the asymptotic variances, compared with the maximum likelihood estimator, which should be acceptable for the more sensible choices of a.

5.3 Convergence of the Empirical Process

To prove (5.2.3), an approach close to that in Randles (1982) can be used. Although his results are not directly applicable, they can be modified.

The main preliminary task is to choose a suitable psuedo metric, d say, on $\{F\}$, the space of distribution functions on S_3, which satisfies

$$n^{\frac{1}{2}}d(\tilde{F}_n, F) = O_p(1) \quad \text{as} \quad n \to \infty \tag{5.3.1}$$

TABLE 5.1

The maximum and minimum values, GES_+ and GES_- respectively, of the influence curves scaled by κ, and the asymptotic efficiencies of the estimators $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$, for different values of a .

	a =	0.1	0.15	0.2	0.25	0.3	0.35
	GES_+/κ	1.34	1.50	1.67	1.86	2.07	2.30
$\tilde{\kappa}_1$	GES_-/κ	-2.54	-1.85	-1.69	-1.58	-1.49	-1.42
	AE	0.85	0.78	0.72	0.67	0.61	0.56
	GES_+/κ	0.43	0.53	0.62	0.72	0.83	0.95
$\tilde{\kappa}_2$	GES_-/κ	-3.91	-2.99	-2.49	-2.16	-1.94	-1.77
	AE	0.59	0.64	0.65	0.64	0.62	0.59
	GES_+/κ	1.11	1.18	1.25	1.33	1.43	1.54
$\tilde{\kappa}_3$	GES_-/κ	-2.56	-2.32	-2.01	-1.85	-1.72	-1.62
	AE	0.90	0.85	0.80	0.75	0.70	0.65

where F is the underlying distribution and

$\tilde{F}_n = n^{-1} \sum_{i=1}^{n} \delta_{x_i}(.)$ is the empirical distribution function for a random sample $x_1, \ldots x_n$ from F. The natural (psuedo) metric for our problem is based on the spherical caps, the collection of sets of the form

$$A(\gamma,t) = \{x \varepsilon S_3 : x'\gamma \geq t\} \quad \gamma \varepsilon S_3, t \varepsilon [-1,1] . \qquad (5.3.2)$$

This metric is defined by

$$d(F,G) = \sup_{\gamma,t} |\text{prob}_F\{A(\gamma,t)\} - \text{prob}_G\{A(\gamma,t)\}| \qquad F,G \varepsilon \{F\} \qquad (5.3.3)$$

or, in the notation of section 5.2,

$$d(F,G) = \sup_{\gamma,t} |F_\gamma(t) - G_\gamma(t)| \qquad F,G \varepsilon \{F\} \qquad (5.3.4)$$

That d is a psuedo metric is obvious. Also, $d(F,G) = 0$ implies that $F = G$, since if $d(F,G) = 0$, F and G agree on $\{A\}$, the set of spherical caps, and therefore (since this is a probability space) on the completed σ-field generated by $\{A\}$, which is precisely the usual Lebesgue σ-field. Hence d is a metric.

Note: in certain simple cases this distance between two distributions can be calculated. If F_1 and F_2 are Fisher distributions with common mean direction μ and concentrations κ_1 and κ_2 respectively, then if $\kappa_1 > \kappa_2$

$$d(F_1,F_2) = (e^{\kappa_1} - e^{\kappa_1 \zeta})/(e^{\kappa_1} - e^{-\kappa_1}) - (e^{\kappa_2} - e^{\kappa_2 \zeta})/(e^{\kappa_2} - e^{-\kappa_2})$$

where $\zeta = [\log\{(\kappa_2 \sinh \kappa_1)/(\kappa_1 \sinh \kappa_2)\}]/(\kappa_1 - \kappa_2)$. However, except in such special cases, these distance calculations cannot be handled analytically.

It will be necessary to consider the empirical process associated with d. The empirical process, indexed by the spherical caps, is defined as

$$\nu_n(A) = n^{\frac{1}{2}}\left[\mathrm{prob}_{\tilde{F}_n}(A) - \mathrm{prob}_F(A)\right] \quad A \varepsilon \{A\} \tag{5.3.5}$$

It is easy to see that

$$E\left[\nu_n(A)\right] = 0 \quad \text{for all} \quad A \varepsilon \{A\}$$

and
$$\mathrm{cov}(\nu_n(A_1), \nu_n(A_2)) = \mathrm{prob}_F(A_1 \cap A_2) - \mathrm{prob}_F(A_1)\mathrm{prob}_F(A_2) \tag{5.3.6}$$

Our aim is to check that (5.3.1) holds, which is equivalent to

$$\sup_{A \varepsilon \{A\}} |\nu_n(A)| = O_p(1) \quad \text{as} \quad n \to \infty \tag{5.3.7}$$

This will follow if statements (D1) and (D2) below are valid:

(D1) The empirical process $\nu_n(A)$ converges in distribution in the supremum norm to a zero-mean Gaussian process, $G_F(A)$ say indexed by the spherical caps, and with covariance function given by (5.3.6). In particular, this implies that

$$\sup_A |\nu_n(A)| \to \sup_A |G_F(A)| \quad \text{in distribution.}$$

(D2) The supremum of the limiting process $G_F(A)$ is bounded in probability, i.e. $\sup_A |G_F(A)| = O_p(1)$.

Certain measure-theoretical difficulties arise in the checking of (D1) (see Billingsley (1968, section 18)), which are bound up with the non-separability of function spaces with the 'uniform' topology. However, a weak convergence theory, described by Dudley (1978), has been designed specifically to cope with such problems. We shall explain how his results can be applied in this case.

The first point to note is that (D1) and (D2) , plus (D3) which is stated below, are tantamount to necessary and sufficient conditions for the set of spherical caps, $\{A\}$, to be a Donsker class (Dudley (1978, p.902)). (NB: the measurability condition referred to by Dudley in his definition of Donsker class is clearly satisfied by $\{A\}$.)

From Corollary 7.18 in Dudley (op. cit.) it follows that $\{A\}$ is, in fact, a universal Donsker class: in particular, (D1) and (D2) are satisfied for any underlying distribution F . Three things need to be checked.

(i) S_3 is a Polish space.

(ii) $\{A\}$ is a Vapnic-Cervonenkis class in S_3 .

(iii) $\{A\}$ is a 'Suslin measurable collection of closed sets for the Effros Borel structure'.

A Polish space is a topological space metrizable by a complete separable metric (Dudley (op. cit., p.909)). So clearly S_3 is Polish, and (i) is satisfied. That (ii) is also satisfied follows immediately from the definition of Vapnic-Cervonenkis class (see Dudley (op. cit., p.920); the Vapnic-Cervonenkis number, in his notation $V(\{A\})$, is 5). Condition (ii), in effect, means that $\{A\}$ is not 'too large'.

A set in a metric space is, by definition, Suslin if it is the range of a continuous function on a Polish space (Dudley (op. cit., p.909)). Since we can parameterise the spherical caps by $(\gamma, t) \varepsilon S_3 \times [-1,1]$, it follows that $\{A\}$ is Suslin, and a 'measurable collection of closed sets'. Then Corollary 7.18 follows directly from Propositions 3.2 and 4.3 and Theorem 7.1 in Dudley (op. cit.).

To summarise: the statements (5.3.7), and therefore (5.3.1), are true for any underlying distribution F on S_3.

The third characteristic of a Donsker class may be stated as follows:

(D3) The sample paths of the limiting process $G_F(A)$ (have versions which) are uniformly continuous with probability 1. (This is the '$G_P UC$' property in Dudley's terminology.)

Continuity is with respect to the psuedo metric m_F on $\{A\}$ defined by

$$m_F(A_1,A_2) = \text{prob}_F(A_1\backslash A_2) + \text{prob}_F(A_2\backslash A_1) \qquad A_1,A_2 \in \{A\}$$

Of course, it is also possible to define a fixed metric m_o on $\{A\}$ by

$$m_o(A(\gamma_1,t_1),A(\gamma_2,t_2)) = \|(\gamma_1-\gamma_2,t_1-t_2)\| \tag{5.3.8}$$

where $\|\cdot\|$ is the Euclidean metric on R^4. When F is absolutely continuous, the metrics m_o and m_F are equivalent; but if F has atomic points they lead to different topologies on $\{A\}$. Since we shall only be considering absolutely continuous distributions F in section 5.4, this difference need not concern us.

In the next section, we shall need to make use of the fact that $\nu_n(A)$ is asymptotically uniformly continuous, i.e. given $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon)$ such that

$$\text{prob}_F\left[\sup_{A_1,A_2} |\nu_n(A_1)-\nu_n(A_2)| \geq \varepsilon : A_1,A_2 \in \{A\}, m_o(A_1,A_2)<\delta\right] \to 0 \tag{5.3.9}$$

as $n \to \infty$. This follows from Theorem 2 in Wichura (1970), which implies that for almost all fixed 'outcomes' ω, $\nu_n(A)(\omega) \to \nu_\infty(A)(\omega)$ uniformly over $A \in \{A\}$ as $n \to \infty$; plus property (D3).

It should be clear from the discussion in this section that the circular arcs in S_2, the hyperspherical caps in S_p, $p > 3$, and the closed balls in R^q are all universal Donsker classes, so that in particular (5.3.1) and (5.3.7) hold in these cases, with no restriction on the underlying distribution F.

5.4 The Asymptotic behaviour of L-statistic Estimators of Concentration

In this section it is shown that, when the underlying distribution F has oval symmetry about μ, (5.2.3) is true if some mild conditions are satisfied by J, F and $\tilde{\mu}_n$, the estimator of μ. These sufficient conditions may be stated as follows.

(J1) The function J in (5.2.1) is non-negative, and piecewise continuously differentiable with at most a finite number of points of discontinuity, the jumps being bounded both in J and J'.

(J2) J is a non-negative linear combination of a finite number of delta functions,

$$J(.) = \Sigma\lambda_j\delta_{a_j}(.) \qquad \lambda_j \geq 0 \ , \ a_j \in (0,1)$$

(F1) F has a continuous density f which is bounded away from zero.

(F2) F has a continuously differentiable density f. In other words, the tangent plane to f at every point $x \in S_3$ is well-defined, and these tangent planes vary continuously as x varies.

(M) The estimator $\tilde{\mu}_n$ is $n^{\frac{1}{2}}$ - consistent, i.e.

$$n^{\frac{1}{2}}\cos^{-1}(\mu_n{}'\mu) = O_p(1) \quad \text{as} \quad n\to\infty \ .$$

(F1) implies that, in the notation of (5.2.1)

(i) $F_\gamma(t)$ and $f_\gamma(t) = \partial F_\gamma(t)/\partial t$ are jointly continuous in γ and t, and the latter is bounded away from zero, i.e.

$$\inf_{\gamma, t} f_\gamma(t) > 0 \tag{5.4.1}$$

(F2) implies that:

(ii) $F_\gamma(t), f_\gamma(t)$ and $f_\gamma{}'(t) = \partial f_\gamma(t)/\partial t$ are jointly continuous in γ and t, and therefore bounded since they are defined on compact sets. (5.4.2)

(iii) If F has oval symmetry about $\mu \in S_3$ then

$$\lim_{\gamma\to\mu} |T(F,\gamma) - T(F,\mu)| = o(\cos^{-1}(\gamma'\mu)) \quad \gamma \in S_3 \tag{5.4.3}$$

It is in principle a very straightforward matter to prove (5.4.1) - (5.4.3), though the details tend to be a little messy.

Our strategy is as follows: firstly, we shall show that either conditions (J1) and (F1), or (J2), (F1) and (F2), ensure that an expansion of the form

$$T(\tilde{F}_n,\gamma)-T(F,\gamma) = S(F,\gamma;\tilde{F}_n-F)+R_n(\tilde{F}_n,F,\gamma) \tag{5.4.4}$$

is valid where the functional S, which will be given explicitly in each case, is linear in $(\tilde{F}_n-F)$, and plays the role of a derivative; and in addition,

$$\sup_{\gamma\in S_3} |R_n(\tilde{F}_n,F,\gamma)| \quad \text{is} \quad o_p(n^{-\frac{1}{2}}) .$$

Now consider the identity

$$n^{\frac{1}{2}}[T(\tilde{F}_n,\tilde{\mu}_n) - T(\tilde{F}_n,\mu)]$$

$$= n^{\frac{1}{2}}[T(F,\tilde{\mu}_n)-T(F,\mu)] + n^{\frac{1}{2}}[S(F,\tilde{\mu}_n;\tilde{F}_n-F)-S(F,\mu;\tilde{F}_n-F)]$$

$$+ n^{\frac{1}{2}}[R_n(F,\tilde{F}_n,\tilde{\mu}_n) - R_n(F,\tilde{F}_n,\mu)] \qquad (5.4.5)$$

Then, if F has oval symmetry about μ, all three terms on the RHS of (5.4.5) will converge to zero in probability: the first because of (5.4.3) and (M) ; the third, because of the validity of the expansion (5.4.4); and the second because (F1) and (F2) are sufficient to ensure that we can apply Theorem 2.8 in Randles (1982), about which further details will be given later. This will be sufficient to ensure that (5.2.3) holds.

Theorem (5.4.6) If J satisfies (J1) and F satisfies (F1) then $T(F_n,\gamma)$ has an expansion of the form (5.4.4) with differential

$$S(F,\gamma;\tilde{F}_n-F) = -\int_{-1}^{1} (\tilde{F}_{n,\gamma}(t)-F_\gamma(t))J(F_\gamma(t))dt \qquad (5.4.7)$$

Proof Without loss of generality, assume that J (or J') has one point of discontinuity, at $a \in (0,1)$. It will be shown that $T(F,\gamma)$ is uniformly Frechet differentiable (see Randles (1982, section 3) for definition). Namely, that

$$\sup_{\gamma \in S_3} |T(G,\gamma)-T(F,\gamma)-S(F,\gamma;G-F)| = o_p(d(F,G)) \qquad (5.4.8)$$

Then the theorem will follow as an immediate consequence of (5.3.1).

Using integration by parts (as in Boos (1979, Theorem 1)) on $[T(G,\gamma)-T(F,\gamma)]$, it can be seen that

$$T(G,\gamma)-T(F,\gamma) = -\int_{-1}^{1} [K(G_\gamma(t))-K(F_\gamma(t)]\,dt$$

where $K(.)$ is the function $K(y) = \int_0^y J(u)du$. Hence $[T(G,\gamma)-T(F,\gamma)-S(F,\gamma;G-F)]$, where S is given in (5.4.7), is equal to

$$\int_{-1}^{1} [K(G_\gamma(t))-K(F_\gamma(t))-[G_\gamma(t)-F_\gamma(t)]\,J(F_\gamma(t)]\,dt \qquad (5.4.9)$$

Now suppose $d(F,G) < \varepsilon/2$ where ε is small. On the intervals $[-1,F_\gamma^{-1}(a-\varepsilon))$ and $(F_\gamma^{-1}(a+\varepsilon),1]$ we can use Taylor's theorem. Therefore, since (J1) implies that $\sup_{u\varepsilon[0,1]} |J'(u)| = M < \infty$, the absolute value of the integrand in (5.4.9) is bounded above by $d^2(F,G)M/2$. Therefore the absolute value of (5.4.9) is bounded above by

$$(1-(-1))d^2(F,G)M/2 = Md^2(F,G) < M\varepsilon^2/4 \qquad (5.4.10)$$

To bound the integral over $[F_\gamma^{-1}(a-\varepsilon),F_\gamma^{-1}(a+\varepsilon)]$, define

$$V_{G,F,\gamma}(t) = [K(G_\gamma(t))-K(F_\gamma(t))]/[G_\gamma(t)-F_\gamma(t)]-J(F_\gamma(t)) \quad \text{if} \quad G_\gamma(t) \neq F_\gamma(t)$$

$$= 0 \quad \text{if} \quad G_\gamma(t) = F_\gamma(t)$$

Because J is bounded, and K is continuous, it is clear that $V_{G,F,\gamma}(t)$ is bounded over t, by V say. Therefore the integrand in (5.4.9) is bounded by $Vd(F,G)$; and the integral over $[F_\gamma^{-1}(a-\varepsilon), F_\gamma^{-1}(a+\varepsilon)]$ is bounded by

$$Vd(F,G)[F_\gamma^{-1}(a+\varepsilon) - F_\gamma^{-1}(a-\varepsilon)] \ . \qquad (5.4.11)$$

Then (5.4.1) plus an application of the Mean Value Theorem ensure that (5.4.11) is bounded above by

$$V\varepsilon^2/4m \qquad (5.4.12)$$

where $m > 0$ (independent of γ) is a lower bound for $f_\gamma(t)$. Finally, putting (5.4.10) and (5.4.12) together, it is seen that (5.4.8) holds. 000

<u>Theorem (5.4.13)</u> If J satisfies (J2) and F satisfies (F1) and (F2) , then $T(F_n,\gamma)$ has an expansion of the form (5.4.4) with

$$S(F,\gamma;\tilde{F}_n-F) = [a-\tilde{F}_{n,\gamma}(F_\gamma^{-1}(a))][f_\gamma(F_\gamma^{-1}(a))]^{-1} \qquad (5.4.14)$$

where (without loss of generality) $J(.) = \delta_a(.)$ for some $a \in (0,1)$.

<u>Proof</u> It happens that $T(F,\gamma)$ is not uniformly Frechet differentiable in this case. A more direct approach is required. Now when $J = \delta_a$, it is immediate that

$$T(\tilde{F}_n,\gamma) - T(F,\gamma) = \tilde{F}_{n,\gamma}^{-1}(a) - F_\gamma^{-1}(a) \ .$$

Expressing $\tilde{F}_{n,\gamma}^{-1}(a)$ as $F_\gamma^{-1}(a+[F_\gamma(\tilde{F}_{n,\gamma}^{-1}(a))-a])$ and, for each fixed γ using Taylor's Theorem to expand $F_\gamma^{-1}(.)$ about a , it is seen that

$$\tilde{F}_{n,\gamma}^{-1}(a) - F_\gamma^{-1}(a)$$

$$= [F_\gamma(\tilde{F}_{n,\gamma}^{-1}(a))-a][f_\gamma(F_\gamma^{-1}(a))]^{-1}$$

$$- \tfrac{1}{2}[F_\gamma(\tilde{F}_{n,\gamma}^{-1}(a))-a]^2[f_\gamma'(c^*)/f_\gamma^3(c^*)] \qquad (5.4.15)$$

where $c^* = \theta F_\gamma(\tilde{F}_{n,\gamma}^{-1}(a)) + (1-2\theta)a$ and $\theta(\gamma) \in [0,1]$. (F1) and (F2) ensure the validity of this expansion. Since f_γ'/f_γ^3 is bounded above, the theorem will follow from (5.3.1), provided it can be shown that, in probability,

$$n^{\frac{1}{2}}\sup_\gamma|[F_\gamma(\tilde{F}_{n,\gamma}^{-1}(a))-a]-[a-\tilde{F}_{n,\gamma}(F_\gamma^{-1}(a))]| \to 0 \qquad (5.4.16)$$

Define $V_n(\gamma) = n^{\frac{1}{2}}[\tilde{F}_{n,\gamma}^{-1}(a)-F_\gamma^{-1}(a)]$ and $W_n(\gamma) = n^{\frac{1}{2}}[a-\tilde{F}_{n,\gamma}(F_\gamma^{-1}(a))]/f_\gamma(F_\gamma^{-1}(a))$. Dependence of V_n and W_n on a is suppressed, since the latter is fixed. An application of Taylor's Theorem (a second order expansion is required) plus (5.3.1), (5.4.1) and (5.4.2) show that (5.4.16) is equivalent to

$$\sup_\gamma|V_n(\gamma) - W_n(\gamma)| \to 0 \qquad (\Pr^*) \qquad (5.4.17)$$

where $\Pr^*$, the outer probability, is needed because sets involved in the statement of (5.4.17) will not in general be measurable.

Now define $Y_{n,V}(\gamma) = n^{\frac{1}{2}}[F_\gamma(h^*)-\tilde{F}_{n,\gamma}(h^*)]/f_\gamma(F_\gamma^{-1}(a))$ where $h^* = F_\gamma^{-1}(a) + V_n(\gamma)/n^{\frac{1}{2}}$. We need three lemmas.

<u>Lemma (5.4.18)</u> If F satisfies (F1) then for any $k > 0$, $n^{\frac{1}{2}}\sup_{\gamma,t}|\tilde{F}_{n,\gamma}(t) - F_\gamma(t)| \leqslant k$ implies that

$\sup_{\gamma} |V_n(\gamma)| \leqslant (k+n^{-\frac{1}{2}})/m$ where $m = \inf_{\gamma,t} f_\gamma(t)$.

Proof Put $t = \tilde{F}^{-1}_{n,\gamma}(a)$ and for each fixed γ use Taylor's Theorem to expand $F_\gamma(t)$ about $F_\gamma^{-1}(a)$, noting that $|\tilde{F}_{n,\gamma}(\tilde{F}^{-1}_{n,\gamma}(a)) - a| < n^{-1}$ 000

Lemma (5.4.19) If F satisfies (F1) and (F2) , then $|V_n(\gamma) - Y_{n,V}(\gamma)| \leqslant [|V_n(\gamma)|^2 M + 1]/(n^{\frac{1}{2}}m)$ where $M = \sup_{\gamma,t}|f'_\gamma(t)|$ and m is as above.

Proof Again, this involves a straightforward application of Taylor's Theorem. A second-order expansion is required here. 000

Lemma (5.4.20) If F satisfies (F1) and (F2) then $|Y_{n,V}(\gamma) - W_n(\gamma)| \to 0$ (Pr*) .

Proof $Y_{n,V}(\gamma)$ and $W_n(\gamma)$ can be expressed in the form $\nu_n(A(\gamma,t_1))$, $\nu_n(A(\gamma,t_2))$ respectively where $t_2 = F_\gamma^{-1}(a)$ and $t_1 = t_2 + V_n(\gamma)/n^{\frac{1}{2}}$. It follows from lemma (5.4.18) and (5.3.1) that

$$\sup_{\gamma} m_o(A(\gamma,t_1) , A(\gamma,t_2)) \to 0 \qquad (\mathrm{Pr}^*)$$

because then $|t_1 - t_2| = |V_n(\gamma)|/n^{\frac{1}{2}} \to 0$ (Pr*) uniformly in γ . Here, m_o is the metric on $\{A\}$ defined in (5.3.8). 000

Now consider the inequality

$$\Pr{}^*(\sup_{\gamma}|V_n(\gamma) - W_n(\gamma)|)$$

$$\leqslant \Pr{}^*(\sup_{\gamma}|V_n(\gamma)-Y_{n,v}(\gamma)|)+\Pr{}^*(\sup_{\gamma}|Y_{n,v}(\gamma)-W_n(\gamma)|) \tag{5.4.21}$$

That the first term on the RHS of (5.4.21) becomes small for n sufficiently large follows from lemmas (5.4.18), (5.4.19), plus (5.3.1). That the second term becomes small follows from lemma (5.4.20). 000

Finally, we need to check that the second term on the RHS of (5.4.5) converges to zero in probability. Observe that (F1) implies that both $\sup_t|F_\gamma(t) - F_\mu(t)|$ and $\sup_t|F_\gamma^{-1}(t) - F_\mu^{-1}(t)|$ are $o(\cos^{-1}(\gamma'\mu))$ as $\gamma \to \mu$; and (F2) implies that $\sup_t|f_\gamma(t) - f_\mu(t)|$ is $o(\cos^{-1}(\gamma'\mu))$ (the details are straightforward). With these results, it is then easy to show that the conditions for Theorem 2.8 in Randles (1982) to apply are satisfied.

We now make some comments.

(i) Since the functional T is linear in J , our results also apply to linear combinations of the form $\lambda_1J_1 + \lambda_2J_2$ where J_1 and J_2 satisfy (J1) or (J2) . The differential S is then the corresponding linear combination of the differentials of J_1 and J_2 .

(ii) The proof that $\sup_{\gamma}|V_n(\gamma) - W_n(\gamma)| \to 0$ (Pr*) in

Theorem (5.4.13) is based on the proof of an analogous (though simpler) result in Ghosh (1971, Theorem 1).

(iii) Theorems (5.4.6) and (5.4.13) apply if S_3 is replaced by S_p , $p = 2$ or $p > 3$; they also apply in R^q if $J : [0,1] \rightarrow R$ has support on $[\psi, 1]$ for some $\psi > 0$.

(iv) Clearly (5.4.6) and (5.4.13) could be proved under weaker assumptions, but such extensions seem to be uninteresting.

5.5 Small Sample Properties

The form of the influence curves of the L-estimators presented in section 5.2 lead one to expect that these estimators will be substantially more robust, in large samples, than the maximum likelihood estimator. As it seems safe to assume that these robustness properties transfer to the small sample case, attention will now be focussed on their small sample properties when the Fisher distribution is in fact correct.

A simulation study was undertaken to investigate the performance of these estimators when the concentration parameter, κ , is at least moderately large. In the simulation procedure, use was made of the fact that if θ is the colatitude, measured from the mean direction, of a unit vector from a Fisher distribution, and ϕ is the longitude, then, for κ at least moderately large

$$z_1 = \sin\theta\cos\phi \quad \text{and} \quad z_2 = \sin\theta\sin\phi \tag{5.5.1}$$

will, effectively, be independent zero-mean Normal variables with variance κ^{-1}.

A large value of κ, $\kappa_0 = 100$, was chosen for convenience. For sample sizes $n = 10, 20, 30, 40$ and 50, n pairs of $N(0,0.01)$ variables were generated using routine G05DDF from NAG (1983). For each pair (z_1, z_2) generated, polar coordinates (θ,ϕ) were obtained from (5.5.1); and for each sample, the estimators $\tilde{\kappa}_0$ (= $\hat{\kappa}$, the maximum likelihood estimator), $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$, based on the sample mean direction, were computed. This was repeated 3,000 times for each sample size, and for $a = 0.1, 0.2$ and 0.3; and, in each case, estimates of the variances and mean square errors of $\tilde{\kappa}_0, \tilde{\kappa}_1, \tilde{\kappa}_2$ and $\tilde{\kappa}_3$ were obtained. The estimates of the variances and mean square errors were then normalised by multiplying through by n/κ_0^2 $(= n/10^4)$, so that, as a consequence of (5.5.1), corresponding estimates for other values of κ not too small, say $\kappa > 3$, can be obtained by multiplying by κ^2. The results are presented in Table 5.2.

Crude and somewhat pessimistic calculations suggest that, with high probability, the most inaccurate estimates in Table 5.2 should be within 10% of the true values; but most of the estimates should be rather more accurate than

TABLE 5.2

Normalised estimates of the variances and mean square errors, $n\mathrm{var}(\tilde{\kappa}_j/\kappa_o)$ and $nE(\tilde{\kappa}_j/\kappa_o-1)^2$ respectively, for $j = 0,1,2,3$.

		n =	10	20	30	40	50	∞
	$\tilde{\kappa}_o$	Var	2.24	1.53	1.28	1.22	1.14	1.0
		MSE	2.89	1.77	1.42	1.32	1.21	1.0
	$\tilde{\kappa}_1$	Var	1.00	1.09	1.08	1.12	1.10	1.18
		MSE	1.26	1.26	1.20	1.21	1.19	1.18
a = 0.1	$\tilde{\kappa}_2$	Var	2.42	2.00	1.81	1.85	1.82	1.70
		MDR	2.46	2.02	1.82	1.85	1.83	1.70
	$\tilde{\kappa}_3$	Var	1.13	1.13	1.08	1.10	1.09	1.11
		MSE	1.27	1.23	1.17	1.16	1.14	1.11
	$\tilde{\kappa}_o$	Var	2.35	1.40	1.34	1.15	1.14	1.0
		MSE	2.99	1.67	1.49	1.24	1.21	1.0
	$\tilde{\kappa}_1$	Var	1.23	1.24	1.31	1.26	1.28	1.38
		MSE	1.44	1.36	1.40	1.34	1.36	1.38
a = 0.2	$\tilde{\kappa}_2$	Var	2.20	1.78	1.82	1.66	1.63	1.54
		MSE	2.29	1.80	1.84	1.67	1.64	1.54
	$\tilde{\kappa}_3$	Var	1.38	1.25	1.31	1.22	1.23	1.25
		MSE	1.43	1.29	1.34	1.25	1.25	1.25
	$\tilde{\kappa}_o$	Var	2.53	1.55	1.32	1.20	1.16	1.0
		MSE	3.20	1.81	1.47	1.28	1.23	1.0
	$\tilde{\kappa}_1$	Var	1.73	1.48	1.41	1.52	1.62	1.63
		MSE	1.94	1.63	1.53	1.62	1.69	1.63
a = 0.3	$\tilde{\kappa}_2$	Var	2.81	2.06	1.88	1.66	1.75	1.61
		MSE	2.93	2.12	1.91	1.66	1.75	1.61
	$\tilde{\kappa}_3$	Var	2.02	1.57	1.46	1.40	1.50	1.43
		MSE	2.04	1.59	1.47	1.42	1.51	1.43

this.

As one might expect, the normalised estimates of the variance and mean square error of the maximum likelihood estimator decrease monotonically to their asymptotic value of 1.0. For each fixed a, the same appears to happen with the quantile estimator $\tilde{\kappa}_2$. However, the trimmed and Winsorised estimators, $\tilde{\kappa}_1$ and $\tilde{\kappa}_3$ respectively, appear to behave differently: for each fixed a, they fluctuate, moderately, about their asymptotic values as the sample size increases.

It is interesting to see in Table 5.2 that for sufficiently small sample sizes, $\tilde{\kappa}_1$ and $\tilde{\kappa}_3$ have estimated variance and mean square error which are somewhat smaller than those of the maximum likelihood estimator; though of course the large sample optimality of the maximum likelihood estimator implies that the reverse should be the case in sufficiently large samples. The results in Table 5.2 appear to suggest that, at least for the range of values of a considered, the smaller the value of a, the larger the sample size needs to be for the maximum likelihood estimator to perform better on these two criteria.

There appears to be little to choose between the trimmed and Winsorised estimators, though both seem to perform somewhat better than the quantile estimator. An estimator which, overall, should be as good as any is

$\tilde{\kappa}_1$, when $a = 0.2$.

An attempt was made to improve the performance of these estimators by reducing the bias, in similar fashion to Best and Fisher (1981). The fact that μ , the mean direction, is being estimated does complicate matters, but multipliers (see section 5.2,(i)) can still be obtained, assuming (incorrectly) that the bias of each order statistic due to the estimation of μ is the same.

However, the estimators $\tilde{\kappa}_1$ and $\tilde{\kappa}_3$, in each case considered in Table 5.2, were negatively biased (unlike the maximum likelihood estimator and $\tilde{\kappa}_2$); and although the 'multiplied' estimators usually did have less bias, their mean square errors never decreased noticably, and in some cases even increased. So there seems little to be gained from this approach, at least for the L-estimators we have considered.

5.6 L-estimators in Other Contexts

As mentioned earlier, (5.2.3) has analogues in other situations. In particular, this includes the Von-Mises distribution on the circle; and it would be reasonable to expect that the best of the L-estimators of concentration have favourable properties in this case as well. However, computational matters are a little less straightforward;

essentially because the χ^2_1 and Von-Mises distribution functions are somewhat less tractable than those of the exponential and truncated exponential distributions. The position is similar for the Von-Mises-Fisher distribution in more than three dimensions.

We recall that FB_5, described by Kent (1982) has a density which may be written

$$c^{-1}\exp\{\kappa x'\mu_3 + \beta[(x'\mu_1)^2 - (x'\mu_2)^2]\} ,$$

where μ_1, μ_2 and μ_3 are three mutually orthogonal unit vectors. One would expect the maximum likelihood estimators of the shape parameters, κ and β, to suffer from a lack of robustness similar to that of the maximum likelihood estimator of the concentration parameter of a Fisher distribution. For a random sample $x_1,\ldots x_n$ of unit vectors from FB_5, consider the squares of the scalar products

$$u_1 = (x_1'\mu_1)^2,\ldots\ u_n = (x_n'\mu_1)^2$$

$$v_1 = (x_1'\mu_2)^2,\ldots\ v_n = (x_n'\mu_2)^2 \quad .$$

It is not difficult to show that estimators of κ and β based on $(u_1,\ldots u_n)$ and $(v_1,\ldots v_n)$, with μ_1 and μ_2 replaced by suitable estimates, satisfy a property analogous to (5.2.3). In other words, the asymptotic properties of such estimators of κ and β doe not depend on whether μ_1 and μ_2 are known or estimated, suitable estimates for μ_1 and μ_2 being the appropriate normalised eigenvectors of the matrix of sums and products (Mardia (1972)) for the sample.

As a consequence, one could, in principle, construct robust estimators of κ and β based on linear combinations of the order statistics

$$\sum_{i=1}^{n} a_j u_{(j)} \quad \text{and} \quad \sum_{i=1}^{n} b_j v_{(j)}$$

with suitable choices for $a_1, \ldots a_n$ and $b_1, \ldots b_n$. If the underlying FB_5 distribution were highly concentrated about its mean direction, then $u_1, \ldots u_n$ and $v_1, \ldots v_n$ would be effectively independent χ^2_1 variables, and so the calculations would not be daunting if the J-function were reasonably simple. However, if the underlying distribution were not highly concentrated, the calculations would be prohibitively complicated, and for this reason estimators of κ and β of this type appear unattractive, because one does not want to have to make an a priori assumption of high concentration.

We conclude by briefly mentioning the possibility of using L-estimators for the eigenvalues of the covariance matrix of a multivariate Normal distribution. For Normal distributions whose covariance matrices have no special structure, one could follow a broadly similar approach to that suggested for FB_5 . In other words, one could project the observations onto estimates of the canonical axes, and use linear combinations of the squares of the projected values to estimate the corresponding eigenvalues. The squares of the projected values would, asymptotically, be independent with χ^2_1 distribution, because elliptical symmetry, plus some mild regularity conditions, are sufficient to ensure that a result analogous to (5.2.3) holds in this case as well.

However, it is important that the estimators of the canonical axes be robust. For situations in which the component Normal variables were, in some sense, measured on compatible scales, a sensible choice of estimator for the mean would be the spatial median, described by Brown (1983), if the ratio of the smallest eigenvalue of the covariance matrix to the largest were not too close to zero. Then one could estimate the orientations of the canonical axes using the normalised eigenvectors of the matrix of sums and products, referred to above, but formed from the projections of the observations onto the hypersphere with centre the estimate of the mean, and radius 1. These estimators of the orientations would be consistent, asymptotically Normal, given a suitable parameterisation, and should be robust against extreme observations. The computation of L-estimators of the eigenvalues of the covariance matrix, given estimates of the canonical axes, is particularly straightforward, because function inversion is not required.

The estimators of the mean and covariance matrix of a Normal distribution described here are translation and rotation invariant, but not affine invariant, unlike the M-estimators suggested by Maronna (1976) and Campbell (1980).

Nevertheless, it would be interesting to know how the estimators described here compare with these M-estimators in situations in which affine invariance is not a strict requirement.

CHAPTER 6

FURTHER RESEARCH

6.1 Rotationally Symmetric Distributions

The Fisher distribution is, for statistical purposes, the most commonly used spherical model. In some situations encountered in practice, it may be reasonable to suppose that the population distribution is unimodal and rotationally symmetric; however, the particular choice of distributional form is usually one of practical convenience. Specifically, the maximum likelihood estimates of the parameters have explicit form, and methods of inference based on the Fisher distribution are quite well developed (see Mardia (1972)). The same comments apply to the von Mises distribution on the circle.

An obvious but important question is: how restrictive is the assumption of a Fisher distribution? Before we can discuss this, it is necessary to specify an alternative class of distributions. If, in this alternative class, we include distributions which do not have rotational symmetry, or even distributions which have rotational symmetry but are not unimodal, then the answer is likely to be: very restrictive. If, however, we define the alternative class to be precisely the unimodal rotationally symmetric distributions, the answer is not so clear. In particular, it is not obvious how to define a measure of 'restrictiveness' in a fruitful way.

One approach would be to find the distribution (in the class of unimodal rotationally symmetric distributions) with given first moment which maximises the distance (according to some suitably chosen metric) to the Fisher distribution with the same first moment. Expressed mathematically, the problem is:

$$\max_{g} d(f,g) \quad \text{where} \quad f(u) = [\kappa/(2\sinh\kappa)]e^{\kappa u} \;,$$

$$g : [-1,1] \to R \;; \qquad (6.1.1)$$

$$g \geq 0 \;,\; g' \geq 0 \;,\; \int_{-1}^{1} g(u)du = 1 \;,\; \int_{-1}^{1} ug(u)du = t \;(= B(\kappa))$$

where d is a metric, perhaps related to the metric defined in Chapter 5, in which case bounds on the error in probability statements would result. Problem (6.1.1) appears not to have an analytic solution, in view of the constraints $g \geq 0$, $g' \geq 0$. However, it may be possible to obtain good approximations to the solution of (6.1.1) numerically, though exactly how one would do this is not clear.

There are other possible approaches to measuring the 'restrictiveness' of the Fisher distribution. For example, something along the lines of Efron and Olshen (1978), who consider the broadness of the class of Normal scale mixtures. Or, alternatively, it might be helpful to express the problem as one in discrimination, in which case the ideas in Collett and Lewis (1981), who consider the problem of discriminating between the von Mises and wrapped Normal distributions, may prove useful.

It may well turn out that in some contexts, particularly those in which very large samples are available, the Fisher family is inadequate, even within the class of unimodal, rotationally symmetric distributions. Solomon and Stephens, in the discussion of Kendall's (1974) paper, express the opinion that (on the circle) it would be helpful to have access to distributions which are more (and less) 'peaked' than the Von-Mises distribution, and suggest the need for a rotationally symmetric family with 2 shape parameters. If this applies on the circle, one would also expect it to apply on the sphere. One such generalisation, FB_4 , was referred to in Chapters 2 and 4. There are other ad hoc generalisations which we have looked at in some detail (but which have not been referred to in earlier Chapters). The FB_4 family seems to be the most attractive generalisation, but it would perhaps be worth mentioning other possibilities briefly.

If θ is the colatitude of a Fisher distribution, measured from the mean direction, then $u_- = (1-\cos\theta)/2$ has negative exponential distribution, truncated at $u_- = 1$. Two obvious possibilities are to generalise from truncated negative exponential on $[0,1]$ to truncated Gamma and Weibull distributions on $[0,1]$. One can do something similar with $u_+ = (1+\cos\theta)/2$. These families are not without interest, but several difficulties do arise. One of these is that the 4-parameter likelihood function (2 shape parameters and 2 location parameters) is unbounded above. The problem is similar to that which arises in maximum likelihood estimation for the 3-parameter Gamma and Weibull distributions.

Another possibility is the Fisher 'mean-mixture'. If the mean direction, μ , of the Fisher distribution $F(\mu,\kappa)$ is itself assumed to have Fisher distribution $F(\mu_o,\kappa_o)$, then the 'mean-mixture' has density

$$g(x|\kappa,\kappa_o,\mu_o) = p(t)/[p(\kappa)p(\kappa_o)] \qquad x \in S_3 \qquad (6.1.2)$$

where $p(.) = [4\pi\sinh(.)]/(.)$ and $t = (\kappa^2+\kappa_o^2+2\kappa\kappa_o x'\mu_o)^{\frac{1}{2}}$. At first (naive) glance, the likelihood function based on this family looks as though, while complicated, it should be well-behaved. However, empirical evidence suggests that the maximum always occurs at a boundary rather than a stationary point. Two cases occur: (i) $\hat{\kappa}_o = \hat{\kappa}$ and (ii) $\hat{\kappa}_o = \infty$ (or $\hat{\kappa} = \infty$). In view of the properties of the analogous Normal mixture, this is not at all surprising. T. Lewis, in some unpublished notes, has provided approximations which illustrate clearly what is happening. This 'degeneracy' in the likelihood problem seems a serious drawback, and it indicates that the mixture is not much different from a Fisher distribution.

So FB_4 does appear to be the most attractive 2-parameter rotationally symmetric generalisation of the Fisher distribution. As in the Fisher case, it would be useful to have an idea of how 'general' this family is.

There is an interesting generalisation of the 'mean-mixture'. If $x_1,\dots x_r$ are independent $F(\mu,\kappa)$ and μ is $F(\mu_o,\kappa_o)$, then the joint marginal distribution of $x_1,\dots x_r$ is:

$$g(x_1,\dots x_r|\kappa,\kappa_o,\mu_o) = p(t)/[[p(\kappa)]^r p(\kappa_o)] \qquad (6.1.3)$$

where $p(.)$ is the same as in (6.1.2), and

$$t = (\kappa_o^2+r\kappa^2+2\kappa\kappa_o\mu_o'\sum_{j=1}^{r} x_j+2\kappa^2\sum_{i<j} x_i'x_j)^{\frac{1}{2}} \quad .$$

When $r>1$, degeneracy in the likelihood problem does not occur. This model is closely related to the 'two-tier' model of McFadden (1982); the latter appears to be an approximation of (6.1.3) based on the assumption that κ and κ_o are large. McFadden's two-tier analysis allows the separation of 'between site' variation (determined by κ_o) and 'within site' variation (determined by κ), and may be regarded as a simple 'random effects' model.

Mixtures of this type and others (e.g. concentration mixtures) have obvious modelling possibilities, and deserve further investigation. Much of the discussion in this section also applies to the von Mises distribution on the circle.

6.2 Spatial Unit Vector Processes

This is an area in which little or no work appears to have been done. We shall just offer a few naive observations, and mention some unsolved problems.

Let S be an index set, assumed to be a subset of p-dimensional Euclidean space. For example, S may be a lattice, or the unit hypercube. In some applications, one may wish to consider unit vector processes defined on S, i.e. a family of (dependent) unit vectors

$$\left\{ x(s) : x(s) \in S_n , \; s \in S \subseteq R^p \right\} .$$

An interesting question is: what is a suitable unit vector analogue of a Gaussian process on S ?

If $n = 2$, so that each $x(s)$ is a circular variable, then there is a straightforward answer. For any real-valued Gaussian process $z(s)$ defined on S, there is an associated circular process obtained by 'wrapping' at each $s \in S$ given by $x(s) = z(s) \pmod{2\pi}$. Several properties of the wrapped process follow immediately.

(i) If the original process has continuous or differentiable sample functions, then so does the wrapped process.

(ii) The marginal distribution of each $x(s)$ is wrapped Normal.

(iii) The finite dimensional distributions can be written down as infinite sums, in similar fashion to the wrapped

Normal.

(iv) The conditional distributions are not of wrapped Normal form.

When $n = 3$, so that each $x(s)$ is spherical, a wrapping procedure is not available. The problem is now much more difficult, and so far we have not been able to make substantial progress. In fact, we have not even succeeded in constructing a joint distribution for $x(s_1)$ and $x(s_2)$ such that the marginal distributions of $x(s_1)$ and $x(s_2)$ are of spherical Brownian motion form (as defined in Roberts and Ursell (1960)). It would be nice to see a solution to this problem.

There is, however, a unit vector process on S which is in principle easily constructed in all dimensions. In the three dimensional case: let $z_1(s)$, $z_2(s)$ and $z_3(s)$ be Gaussian processes defined on S. Then define the unit vector process

$$\begin{aligned} x(s) &= (z_1(s), z_2(s), z_3(s))/R(s) \\ R(s) &= ([z_1(s)]^2 + [z_2(s)]^2 + [z_3(s)]^2)^{\frac{1}{2}} . \end{aligned} \qquad (6.2.1)$$

If S contains an open set, the question arises as to whether $\inf_{s \in S} R(s) > 0$ almost surely. We do not know the answer. If the answer is yes, the sample functions will be continuous or differentiable if those of $z_1(s)$, $z_2(s)$ and $z_3(s)$ are; but if not, the unit vector process will have singularities. This problem can not arise if S is countable.

If $z_1(s)$, $z_2(s)$ and $z_3(s)$ are independent, homogeneous processes with zero mean and the same covariance function, then the finite dimensional distributions simplify considerably.

It would be interesting to know whether the process $x(s)$ constructed via (6.2.1) is infinitely divisible, when a suitable extension of Kent's definition of spherical addition is used (see Mardia (1975, discussion)).

There are many interesting and challenging problems in this and related areas such as 'spatial orientation processes'. Kingman (1984) mentions a problem concerning the orientation of grains in a crystalline substance.

REFERENCES

Amos, D.E. (1974). Computation of the modified Bessel functions and their ratios. Math. of Computation, 28, 239-251.

Barndorf-Nielson, O. (1978). Information and Exponential Families in Statistical Theory. Chichester: Wiley.

Barndorf-Nielson, O. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). J.R. Statist. Soc. B, 41, 279-312.

Beran, R. (1979). Exponential models for directional data. Ann. Statist., 7, 1162-1178.

Best, D.J. and Fisher, N.I. (1979). Efficient simulation of the von Mises distribution. Appl. Statist., 28, 152-157.

Best, D.J. and Fisher, N.I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. Commun. Statist., B10, 493-502.

Billingsley, P. (1968). Convergence of Probability Measures. New York: Wiley.

Bingham, C. and Mardia, K.V. (1978). A small circle distribution on the sphere. Biometrika, 65, 379-389.

Boos, D.D. (1979). A differential for L-statistics. Ann. Statist., 7, 955-959.

Brown, B.M. (1983). Statistical uses of the spatial median. J.R. Statist. Soc. B, 45, 25-30.

Campbell, N.A. (1980). Robust procedures in multivariate analysis. I. Robust covariance estimation. Appl. Statist., 29, 231-237.

Collett, D. (1978). Some problems related to outliers on the line and on the circle. Unpublished Ph.D. thesis, University of Hull.

Collett, D. and Lewis, T. (1981). Discriminating between the von Mises and wrapped Normal distributions. Austral. J. Statist., 23, 73-79.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. London: Chapman and Hall.

De Waal, D.J. (1979). On the normalizing constant for the Bingham-von Mises-Fisher matrix distribution. South African Statist. J., 13, 103-112.

Dickson, L.E. (1945). New First Course in the Theory of Equations. Wiley.

Dudley, R.M. (1978). Central limit theorems for empirical measures. Ann. Probability, 6, 899-929.

Efron, B. (1978). The geometry of exponential families. Ann. Statist., 6, 362-376.

Efron, B. and Olshen, R.A. (1978). How broad is the class of Normal scale mixtures? Ann. Statist., 6, 1159-1164.

Fisher, N.I. (1982). Robust estimation of the concentration parameter of Fisher's distribution on the sphere. Appl. Statist., 31, 152-154.

Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist., 42, 1957-1961.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. J. Am. Statist. Assoc., 69, 383-393.

Kendall, D.G. (1974). Pole seeking Brownian motion and bird navigation (with discussion). J.R. Statist. Soc. B, 36, 365-417.

Kendall, M.G. and Stuart, A. (1963). The Advanced Theory of Statistics. Vol. 1. London: Griffin.

Kent, J.T. (1982). The Fisher-Bingham distribution on the sphere. J.R. Statist. Soc. B, 44, 71-80.

Kingman, J.F.C. (1984). Probability and random processes (with discussion). J.R. Statist. Soc. A, 147, 233-244.

Lewis, T. and Fisher, N.I. (1982). Graphical methods for investigating the fit of a Fisher distribution to spherical data. Geophys. J.R. Astr. Soc., 69, 1-13.

McFadden, P.L. (1980). The best estimate of Fisher's parameter κ. Geophys. J.R. Astr. Soc., 60, 397-407.

Hartmann, P. and Watson, G.S. (1974). "Normal" distribution functions on spheres and the modified Bessel functions. Ann. Probability, 2, 593-607.

McFadden, P.L. (1982). Two-tier analysis in paleomagnetism. Geophys. J.R. Astr. Soc., 71, 519-543.

Mardia, K.V. (1972). Statistics of Directional Data. London: Academic Press.

Mardia, K.V. (1975). Statistics of directional data (with discussion). J.R. Statist. Soc. B, 37, 349-393.

Mardia, K.V. and Spurr, B.D. (1973). Multi-sample tests for multi-modal and axial populations. J.R. Statist. Soc. B, 35, 422-436.

Mardia, K.V. and Sutton, T.W. (1975). On the modes of a mixture of two von Mises distributions. Biometrika, 62, 699-701.

Mardia, K.V., Holmes, D. and Kent, J.T. (1984). A goodness-of-fit test for the von Mises-Fisher distribution. J.R. Statist. Soc. B, 46, 72-78.

Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. Ann. Statist., 4, 51-67.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. Ann. Statist., 10, 65-80.

NAG (1983). Numerical Algorithms Group Fortran Library, Oxford.

Randles, R.H. (1982). On the asymptotic normality of statistics with estimated parameters. Ann. Statist., 10, 462-474.

Roberts, P.H. and Ursell, H.D. (1960). Random walk on a sphere and on a Riemannian manifold. Phil. Trans. Roy. Soc., A252, 317-356.

Schmidt, P.W. (1976). The non-uniqueness of the Australian Mesozoic paleomagnetic pole position. Geophys. J.R. Astr. Soc., 47, 285-300.

Schmidt, P.W. and McDougall, I. (1977). Paleomagnetic and Potassium-Argon dating studies of the Tasmanian Dolerites. J. Geol. Soc. of Australia, 25 (Part 6), 321-328.

Schou, G. (1978). Estimation of the concentration parameter in von Mises-Fisher distributions. Biometrika, 65, 369-377.

Selby, B. (1964). Girdle distributions on the sphere. Biometrika, 51, 381-392.

Serfling, R.J. (1984). Generalised L-, M- and R-statistics. Ann. Statist., 12, 76-85.

Ulrich, G. (1984). Computer generation of distributions on the m-sphere. Appl. Statist., 33, 158-163.

Watson, G.S. (1973). Review of 'Statistics of Directional Data' by K.V. Mardia. Technometrics, 15, 935-936.

Watson, G.S. and Irving, E. (1957). Statistical methods in rock magnetism. Mon. Not. R. Astr. Soc., Geophys. Suppl., 7, 289-300.

Wichura, M.J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent image laws. Ann. Math. Statist., 41, 284-291.

Wood, A. (1982). A bimodal distribution on the sphere. Appl. Statist., 31, 52-58.

Wood, A. (1984). Simulation of Spherical Distributions in the Fisher-Bingham family. Unpublished first version.

Yfantis, E.A. and Borgman, L.E. (1982). An extension to the von Mises distribution. Commun. Statist., All, 1695-1706.